Developmental Review 37 (2015) 66–108

Contents lists available at ScienceDirect

Developmental Review

journal homepage: www.elsevier.com/locate/dr

Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition

Lucy C. Erickson *, Erik D. Thiessen

Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA

ARTICLE INFO

Article history:
Received 19 June 2014
Revised 20 May 2015
Available online 23 June 2015

Keywords:
Statistical learning
Language acquisition
Individual differences
Developmental disorders

ABSTRACT

Considerable research indicates that learners are sensitive to probabilistic structure in laboratory studies of artificial language learning. However, the artificial and simplified nature of the stimuli used in the pioneering work on the acquisition of statistical regularities has raised doubts about the scalability of such learning to the complexity of natural language input. In this review, we explore a central prediction of statistical learning accounts of language acquisition – that sensitivity to statistical structure should be linked to real language processes – via an examination of: (1) recent studies that have increased the ecological validity of the stimuli; (2) studies that suggest statistical segmentation produces representations that share properties with real words; (3) correlations between individual variability in statistical learning ability and individual variability in language outcomes; and (4) atypicalities in statistical learning in clinical populations characterized by language delays or deficits.

Introduction

To acquire language, infants must learn a vast number of individual words, expressions, and grammatical constructions. The speed with which infants succeed at this immense undertaking has impressed theorists for many years (e.g., Chomsky, 1959). With respect to vocabulary alone, a typical university

* Corresponding author. Department of Psychology, Carnegie Mellon University, Baker Hall 342C, Pittsburgh, PA 15213, USA. Fax: +1 412 286 2798.
E-mail address: lerickso@andrew.cmu.edu (L.C. Erickson).

http://dx.doi.org/10.1016/j.dr.2015.05.002
0273-2297/

graduate has a receptive vocabulary of 150,000 words (Miller & Gildea, 1987). An average 5th grader knows 40,000 words. By age 6, a typically developing child comprehends 10,000 words (Anglin, Miller, & Wakefield, 1993). Even learning a single word poses multiple learning challenges. At minimum, infants must associate a referent with a word form. To do so, infants must have some knowledge of the word form – some memory representation of the sounds of the word – before they can create a mapping between a word form and its meaning. Discovering the acoustic form of a word is itself problematic because infants hear relatively few words in isolation, even in infant directed speech (Brent & Siskind, 2001; Van de Weijer, 2001). Moreover, simply recognizing a word form does not suffice for word learning, because the word's meaning is only learned when the infant discovers which of the many possible items and events in the environment to which the word form refers. One learning process that may help infants learn the meaning of words, and provide useful information about many other aspects of language, is *statistical learning*.

Statistical learning refers to learning on the basis of some aspect of the statistical structure of elements of the input, primarily their frequency, variability, distribution, and co-occurrence probability. To illustrate how statistical learning might be useful for language acquisition, consider the problem of word learning once again. The co-occurrence of sounds in the input can help infants discover words as a function of the probability with which they occur together (e.g., Newport & Aslin, 2000). For example, in the phrase *happy#doggie*, the syllables within the words *happy* and *doggie* predict each other more reliably than the syllables that span the word boundary, because syllable combinations that occur incidentally between words (e.g., *py#do*; the end of *happy* and the beginning of *doggie*) are less likely to occur than combinations that co-occur within words. Similarly, the co-occurrence of lexical forms and objects or events in the environment can provide infants with information about the referent of a particular lexical form (e.g., Yu & Smith, 2007; Vouloumanos, 2008). For example, a word like *doggie* is more likely to occur in the presence of a canine than in the presence of a fork, information that can help infants pair the lexical form with the appropriate referent.

Statistical learning accounts of many aspects of language acquisition, including word segmentation (e.g., Perruchet & Vinter, 1998; Swingley, 2005; Saffran, Aslin, & Newport, 1996; Thiessen, Kronstein, & Hufnagle, 2013), phonological learning (Maye, Werker, & Gerken, 2002; Thiessen & Saffran, 2003, 2007), and syntactic learning (Thompson & Newport, 2007; Tomasello, 2000), arose from studies that directly manipulated the statistical structure of artificial linguistic input (e.g., Saffran, Aslin et al., 1996). These accounts have been influential for the past 15 years and have generated many productive lines of research (e.g., Newport & Aslin, 2004; Thiessen & Saffran, 2003). Perruchet and Pacton (2006) described the field of statistical learning as "growing exponentially" (p. 233). However, statistical learning approaches rely primarily on research conducted in a laboratory setting with artificial toy languages (e.g., one study familiarized infants with strings generated from a miniature artificial grammar and subsequently tested them on their ability to discriminate novel strings that obeyed the rules of the grammar from illegal strings, which found that children can generalize information from these grammars to novel grammatical strings; Gomez & Gerken, 1999). Criticisms of statistical learning approaches have raised doubts about the ability of such laboratory findings to scale up to the complexity of real language (e.g., Johnson & Seidl, 2009; see also Pierrehumbert, 2003, 2006). The goal of this paper is to explore the feasibility of statistical approaches for the acquisition of natural languages. This will be accomplished by (1) briefly describing a theoretical account of statistical learning through which relevant phenomena will be discussed, (2) discussing the criticisms of some of the early studies as well as some recent studies that have tried to address these criticisms, and (3) exploring predictions statistical learning accounts make about language acquisition. Some of these predictions have already been investigated empirically, whereas others will need to be addressed in future studies.

The extraction and integration framework

The first experiments on infant statistical learning were focused on word segmentation (Saffran, Aslin et al., 1996). Those experiments demonstrated that infants could segment fluent speech on the basis of the co-occurrence probability between adjacent syllables, a statistical feature called "transitional probability" (Aslin, Saffran, & Newport, 1998). Although the term "statistical learning" has been frequently taken to be synonymous with sensitivity to transitional probabilities, subsequent work has demonstrated that language learners are sensitive to a broader set of statistical structures, such as the distribution of exemplars across a continuum (e.g., Maye et al., 2002). To describe the range of statistical structure to which infants are sensitive, and to examine how this sensitivity might contribute to language acquisition, we will rely on the Extraction and Integration Framework (Thiessen et al., 2013). This framework argues that learners are sensitive to two aspects of statistical structure: conditional statistical information and distributional statistical information. Sensitivity to these two aspects of the input, in turn, arises from two complementary processes: extraction and integration. Extraction refers to the process of holding two distinct elements of the input in working memory and binding them together into a single chunk (Perruchet & Vinter, 1998). Integration refers to the process of combining information across stored chunks to identify central tendencies and prototypical information (Hintzman, 1984; Thiessen & Pavlik, 2013). Whereas these two processes are complementary and interactive, they are posited to reflect at least partially independent processes. For example, extraction relies on working memory and is thought to be guided by attention (e.g., Perruchet & Tillmann, 2010; Thiessen et al., 2013).[1] As such, it should be mediated by the frontal brain networks involved in working memory and attention to a greater extent than integration. In contrast, integration may more be reliant on hippocampal and cortical structures associated with long-term memory, in line with the Complementary Learning Systems hypothesis (McClelland, McNaughton, & O'Reilly, 1995).

While we conceptualize extraction and integration as separate processes, it remains to be seen how separable these two processes are. First, it may well be the case that no single statistical learning task is a "pure" measure of either extraction or integration; experimental and naturalistic learning tasks may always invoke both, to greater or lesser degrees. At a minimum, these two processes are highly interactive, as the units that are extracted from perceptual input are subsequently integrated over to identify consistent patterns in the input. Second, it may be the case that extraction and integration share partially overlapping mechanisms, or are even distinct surface realizations of the same underlying computations (e.g., Frost, Armstrong, Siegelman, & Christiansen, 2015; Thiessen & Erickson, 2013b; Thiessen & Pavlik, 2015). For example, a key process in extraction is binding disparate elements of the input into a discrete representation. Although this process is dependent on attention and working memory to select the to-be-bound elements (e.g., Baker, Olson, & Behrmann, 2004), the process of binding may also involve the hippocampus (a structure we have suggested is also involved in integration), given its role in memory formation. Neurological data are at least partially consistent with this account, as at least some extraction tasks have been shown to involve hippocampal activation, which is more consistent with long-term memory processes (Kim, Lewis-Peacock, Norman, & Turk-Browne, 2014; see also Schapiro, Gregory, Landau, McCloskey, & Turk-Browne, 2014 for evidence that medial temporal lobe damage including the hippocampus disrupts statistical segmentation, an extraction task; but see Knowlton, Ramus, & Squire, 1992 for evidence of intact Artificial Grammar Learning in amnesiac patients). Similarly, attentional processes (which we hypothesize to be related to extraction) have been shown to stabilize extracted representations in the hippocampus (Aly & Turk-Browne, 2015)[2].

Despite these uncertainties about the processes underlying statistical learning, the distinction between the processes of extraction and integration (at least descriptively) is a useful one for at least two reasons. First, it highlights the fact that language acquisition involves sensitivity to more kinds of statistical information than simple transitional probabilities. Second, this account gives rise to several novel predictions, as well as empirically testable (and thus falsifiable) claims about how statistical learning contributes to language acquisition. For example, research on the role of attention in statistical learning has yielded inconsistent results. Although it is clear that attention plays a role in at least some forms of statistical learning (e.g., Baker et al., 2004; Toro, Sinnett, & Soto-Faraco, 2005), the way in

[1] Although we use the terms working memory and attention we do not mean to imply that this is necessarily related to voluntary or effortful processes rather than exogenous or stimulus-driven processes. In our account, the term working memory is interchangeable with short-term memory, although we acknowledge that uncertainty exists regarding whether these processes represent unitary or separable capacities (e.g., Unsworth & Engle, 2007), or whether they only operate on consciously represented information (Soto, Mäntylä, & Silvanto, 2011).

[2] It should be mentioned that this research has been conducted with adults for largely practical reasons. An important task for future research will be to explore whether similar support for these proposals will be found with infant populations, given the developmental nature of the phenomena the Extraction and Integration Framework is intended to explain.

which attention interacts with or influences statistical learning is unclear. In some tasks, attention appears to impair statistical learning, whereas in others, attention appears to be helpful (Finn, Lee, Kraus, & Kam, 2014). A two-process account may help clarify these conflicting results. Because extraction is driven, at least in part, by attention (elements of the input are only bound together when they are simultaneously held in working memory, which can be altered by attention), tasks that require extraction may benefit from attention, as it allows learners to bind elements together more quickly or efficiently. By contrast, integration is a more passive process, in which the central tendency of the input emerges over assimilation of information across exemplars. We believe that this process of assimilation occurs automatically, via memory processes related to spreading activation as a function of similarity. To the extent that two exemplars (i.e., memory traces) are similar, they will activate each other (e.g., Hintzman, 1984). When two memory traces are similar and activate each other, the information contained in each is mingled into a representation reflecting the central tendency of both, similar to theoretical accounts of prototype formation (Bomba & Siqueland, 1983; Posner & Keele, 1968; Thiessen & Pavlik, 2013). In integration tasks where participants are not required to bind together disparate elements of the input, attention and working memory may be less necessary, suggesting that these kinds of tasks may be less affected by attentional manipulations. Below, we will flesh out this framework more thoroughly, with an eye toward explaining how statistical learning contributes to language acquisition, and highlight novel predictions of this account for language learning.

Conditional statistical learning

The most well-known finding in the statistical learning literature is that infants can readily segment speech on the basis of its conditional statistical structure (Saffran, Aslin et al., 1996). When exposed to an artificial language input, infants extract syllables that co-occur reliably and appear not to learn syllable groupings that occur together less predictably (Aslin et al., 1998). As these experiments demonstrate, infants (and adults) are sensitive to the conditional likelihood that one event will occur, given information that another event has happened. Traditionally, speech segmentation studies have manipulated one well known type of conditional probability, namely, transitional probability (Harris, 1954; Hayes & Clark, 1970; Saffran, Aslin et al., 1996). Transitional probability is defined as the probability that some event Y will occur given that some other event X has already occurred. It is measured as the number of times that event XY occurs divided by the overall frequency of X. For example, if XY occurs 60 times and X occurs 100 times, the transitional probability of X to Y is 0.6. Thus, transitional probability incorporates raw frequency of co-occurrence but is more robust than mere co-occurrence because items can occur together frequently simply because they are both high frequency items (e.g., *the dog*; Aslin et al., 1998). Statistical approaches to segmentation gained prominence following a pioneering study by Saffran, Aslin et al. (1996). This study was designed to test whether 8-month-old infants were able to extract words from a brief exposure to fluent speech when the only viable cue to word boundary was the presence of higher conditional probabilities between syllables comprising words relative to syllables spanning word boundaries. The authors manipulated conditional probabilities directly using an artificially synthesized stream of words, or tri-syllabic sequences (e.g., *bidaku*, *padoti*). Each syllable within a word predicted the following syllable with 100% reliability. Following a 2 minute exposure to this speech stream, infants were tested on their ability to discriminate words in the language (e.g., *bidaku*) from part–word foils, which were constructed by concatenating the end of one word to the beginning of another (e.g., *kupado*). Infants showed a preference for part–word foils that indicated that they successfully discriminated these two types of test items. The authors interpreted this behavior as evidence that conditional probabilities between syllables may support early infant word segmentation, and coined the phrase "statistical learning" to describe the learning mechanism involved.

Although some computational approaches explicitly calculate transitional probabilities between conditional elements (e.g., Gambell & Yang, 2004), the Extraction and Integration Framework accounts for sensitivity to conditional probabilities based on features of the human memory system rather than transitional probability calculations (e.g., Perruchet & Vinter, 1998; Thiessen et al., 2013). To accomplish this, the framework invokes PARSER, a computational model developed by Perruchet and Vinter (1998). PARSER uses three memory-based processes to extract discrete coherent chunks from the input: activation, decay, and interference. This model works by taking in an input, which consists of a string of syllables, and randomly grouping adjacent syllables from that input into a chunk or a percept. Once a percept is created, it receives a certain amount of activation. Because of decay, this level of activation decreases over time, but can increase if the percept is re-encountered as the model proceeds through the input. Spurious groupings (e.g., syllables that span word boundaries) are less frequent and thus less likely to receive activation. Similarly, spurious groupings will show greater decay as they will be re-encountered less frequently. Finally, interference will also contribute to increasing the activation of coherent items relative to spurious groupings. In the model, chunks which share elements (e.g., syllables) compete with each other. Thus, each chunk created by a spurious grouping across a word boundary will receive interference from two items because both words that form the chunk will provide competition (Perruchet & Vinter, 1998). As a result, coherent items such as words will tend to win out over less frequent and less coherent syllable clusters, and will be more likely to be stored in a lexicon of word forms. In this way, PARSER is able to explain how sensitivity to conditional probabilities rather than mere co-occurrence is achieved. Take as an example the phrase *yellow#ducky.* Although PARSER might initially chunk *lowdu*, repeated presentations of the phrase are more likely to accurate chunk the two words as *yellow* and *ducky.* Because activation for *lowdu* will decay over time unless it is chunked again, and because it will receive interference from both of its constituent words, activation for *yellow* and *ducky* will tend to increase whereas activation for *lowdu* will tend to decrease. Together, these fundamental memory processes are able to explain infant sensitivity to statistical structure based on conditional probabilities without invoking any explicit computations of transitional probabilities.

Distributional statistical learning

The Extraction and Integration Framework explains the extraction in terms of a chunking model such as PARSER to explain the sensitivity to conditional statistics as well as other language-general cues to word boundary. However, the process of chunking fails to explain other phenomena that result in adaptation to environmental statistical regularities such as category learning (e.g., Maye et al., 2002). Thus, integration is required to complement the process of extraction. The process of integration involves combining information across exemplars to form an aggregate representation that captures the central tendency of those exemplars. This process allows for sensitivity to distributional structure, such as the distribution of exemplars along a continuum. Distributional statistical learning is our term for sensitivity to those aspects of the statistical structure of the input that capture the frequency and variability of exemplars in the input.

As an example, consider the case of an infant learning forming two phonetic categories, /*d*/ and (unaspirated) /*t*/, which differ in voice onset time (VOT). Maye et al. (2002) found that the frequency and variability of the exemplars to which infants were exposed influenced whether they formed two categories or a single category that subsumed both /*d*/ and (unaspirated) /*t*/. When infants were presented with a bimodal distribution of sounds that included prototypical exemplars of the two phonemes, infants were more likely to exhibit evidence of discriminating exemplars of the two categories. In contrast, infants who experienced a unimodal distribution that frequently included a sound intermediate between the two prototypical phoneme exemplars were less likely to show evidence of discrimination. Distributional statistical learning has been recently instantiated in a computational model called iMINERVA, using principles of exemplar memory models (Hintzman, 1984; Thiessen & Pavlik, 2013).

In this approach, iMINERVA accounts for various aspects of statistical learning via comparing current and prior exemplars, and creating an integrated representation that incorporates the central tendencies of these stored exemplars. The exemplars are coded as n-dimensional vectors with positive and negative feature valences that are linked to the presence or absence of certain characteristics (e.g., roundness; where a positive value for roundness would indicate the presence of that feature). The magnitude of the features indicates the certainty of the presence or absence of the characteristic, such that larger values indicate greater certainty of the presence of the feature whereas smaller (i.e., negative) values indicate greater certainty of the absence of the feature. Learning is accomplished using memory based principles of similarity-based comparison, decay, integration, and abstraction. In integration, when the similarity of prior experiences (i.e., exemplars) to current information passes a certain threshold, a process is initiated in which prior and current exemplars are integrated to form a new representation. As part of this process, both old and new exemplars are stored along with an interpretation of the experience. This new representation is formed through an additive process in which the feature vectors of current and prior exemplars are merged. If the current exemplar and prior exemplars both have consistent features, the new representation will have more extreme values than either of them. If the current and prior exemplars have conflicting feature values (i.e., have values that go in the opposite directions), the new representation will have less extreme values for the features that it has previously. Depending on a learning rate parameter, the current exemplar can have more or less influence on the new vector formed during the comparison. Finally, an abstraction parameter facilitates generalization to novel stimuli, through the transformation of some features to null values so that they can no longer be used to compute similarity ratings. This process was chosen to simulate the finding that experience is often accompanied by a decrease in sensitivity to certain features of the input (e.g., Werker & Tees, 1984). These four memory-based processes (similarity-based activation, integration, decay, and abstraction) together simulate distributional learning.

The Extraction and Integration Framework suggests that these exemplar memory structures, when coupled to chunking models such as PARSER, are sufficient to explain both the discovery of word forms in speech on the basis of conditional probability information as well as the learning about the central tendency of exemplars (e.g., Maye et al., 2002), the role of variability in the discovery of nonadjacent regularities (e.g., Gómez, 2002), and cue-weighting phenomena (e.g., Thiessen & Erickson, 2013a). As an example, consider how this framework can account for the detection nonadjacent relations. Although nonadjacent relations are more difficult to learn than adjacent regularities, learners are able to detect them (e.g., Creel, Newport, & Aslin, 2004; Gómez, 2002). In some cases, nonadjacent regularities can only be detected if highlighted by a perceptual or structural cue (e.g., Creel et al., 2004). For example, Gómez (2002) found that variability in the intervening element is critical to infants' detection of a nonadjacent variability, such that infants only detected the relationship between *a* and *b* in the sequence *a-X-b* when *X* was drawn from a set of 24 possible *X* elements than when it was drawn from 12 or 3 *X* elements. Although chunking models alone such as PARSER cannot account for the learning of non-adjacent relations because they simply store discrete, unitized representations of adjacent syllables, iMINERVA solves this problem using an abstraction parameter. Because when infants experience *a-X-b* strings with only a few surface instantiations of the *X* element, the entirety of the string is stored including details of the *X* element. In contrast, when the *X* element is drawn on many distinct elements, the specific features of various stored *X* elements cancel out and are abstracted away, leaving only the nonadjacent relation to be detected between *a* and *b*. In this way, iMINERVA accounts for the classic finding that variability in intervening elements allows learners to detect nonadjacent relation.

Word segmentation: an illustration of extraction and integration working in concert

Extraction of words from the speech stream is crucial for language learning, but this process is difficult for several key reasons. Word boundaries in speech are neither marked by pauses in the way that white space separates words in print (Cole & Jakimik, 1980), nor is there one fully reliable cue to indicate where one word ends and the next begins. Instead, word boundaries are marked by a convergence of imperfect cues. As a result, words in speech run together and adults are only able to comprehend speech effortlessly because of the knowledge – of words and of the structure of their native language – that they bring to bear on word segmentation. This segmentation problem can be experienced by any adult learner of a second language, despite the fact that an adult is likely to possess general knowledge about how languages work, as well as meta-cognitive strategies that help parse an unfamiliar language. Parsing speech, then, poses an even greater challenge for infants because they lack both the knowledge and the explicit strategies that may provide scaffolding for adults and older children (e.g., knowing that certain clusters of consonants are unlikely to occur at the beginning of words, or that a stressed syllable is typically a word onset in English; Cutler, 1996; Cutler & Carter, 1987). Consistent with this assertion, many studies indicate that word segmentation is challenging for infants (e.g., Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Johnson & Jusczyk, 2001; Jusczyk & Aslin, 1995).

Despite this inherent challenge, a seminal study by Jusczyk and Aslin (1995) revealed that 7.5-month-old infants familiarized with passages of speech were able to recognize words from those passages, which indicates that the ability to segment speech is present by at least this age. Consequently, when and how infants begin to segment speech from continuous input has generated considerable research over the past 30 years (e.g., Morgan, 1996; Saffran, Newport, & Aslin, 1996; Thiessen, Hill, & Saffran, 2005). This research has generated competing theories about the mechanisms and the cues that allow infants to segment speech successfully (e.g., Johnson & Tyler, 2010; Seidenberg, 1997; Thiessen & Saffran, 2003; Swingley, 2005; Yang, 2004). Many of these approaches can be characterized as acoustic or phonological theories; that is, they suggest that infants identify words in the input on the basis of the typical sound structure of words in the native language. In English, for example, open class words typically begin with a stressed syllable (Cutler & Carter, 1987). Once infants have learned this regularity, they can use lexical stress as a cue to word segmentation (e.g., Johnson & Jusczyk, 2001; Jusczyk, Houston, & Newsome, 1999). However, these kinds of approaches face a chicken and egg problem. Phonological regularities are useful for learning words, but to identify most of these phonological regularities, infants must be familiar with enough word forms to discover the regularities that characterize words in their native language (Thiessen & Saffran, 2003).

Sensitivity to conditional and distributional statistical structure may provide infants with a way to identify phonological regularities that is not dependent on knowledge of native-language phonological regularities. Proponents of statistical learning approaches to word segmentation hold that statistical regularities in the sequences of sounds that comprise words play a critical role in infants' initial word discovery (e.g., Thiessen & Saffran, 2003). The roots of statistical approaches to infant word segmentation originate in an observation made in the 1950s: within speech corpora, the sound sequences that occur within words are more likely to occur together than the sound sequences that occur incidentally across word boundaries (Harris, 1955; Hayes & Clark, 1970). A byproduct of this distributional fact is that statistical clustering of syllables in words can be used to extract word forms (frequent and statistically coherent clusters of syllables) without requiring any language-specific knowledge of the phonological regularities that indicate word boundaries (e.g., knowing that a stressed syllable is likely to mark a word onset in English; Cutler & Carter, 1987). Because statistical clustering does not require adaptation to the phonological structure of the native language, it is flexible enough to explain both how segmentation proceeds in languages that lack rhythmic cues to word boundaries, as well as how infants who have not fully adapted to the structure of their native language initially segment speech. Thus, proponents of statistical approaches argue that sensitivity to these conditional statistical regularities is what allows infants to extract their earliest set of word forms from fluent speech.

The importance of identifying an early proto-lexicon extends beyond learning to recognize the word forms themselves. To the extent that these segmented word forms conform to the patterns of an infant's native language, familiarity with a few (statistically segmented) word forms can provide an opportunity to identify language-specific acoustic cues. This process is an example of distributional statistical learning and involves integrating across stored exemplars (e.g., word forms) to learn something about their central tendency (e.g., Thiessen & Pavlik, 2013). For example, infants may induce phonological patterns shared by a set of learned word forms, such as syllable onsets are voiced and syllable offsets are unvoiced (Saffran & Thiessen, 2003), or that words begin with /t/ (Sahni, Seidenberg, & Saffran, 2010). A corollary of this approach is that statistical regularities serve as one of the earliest cues to word segmentation and are only later supplanted in relative importance by language-specific acoustic cues, which typically do not require multiple repetitions to inform word boundaries (Swingley, 2005; Thiessen & Erickson, 2013b, 2015; Thiessen & Saffran, 2003).

According to the framework we have described, statistical learning is a powerful mechanism available to infants acquiring their native language. Sensitivity to statistical syllable co-occurrence probabilities represents a viable strategy for early infant word discovery. However, word boundaries are also marked by acoustic cues such as lexical stress and phonotactics. To arrive at the most accurate segmentation of speech, a successful learner should integrate the information from these multiple imperfect cues. Thus, accounts of word segmentation must specify how sensitivity to each of these cues arises, and how infants integrate multiple sources of information online in the segmentation of speech. Although many acoustic accounts of word segmentation phenomena conceptualize sensitivity to acoustic and statistical cues as arising from fundamentally distinct processes (e.g., Shukla, Nespor, & Mehler, 2007), the Extraction and Integration Framework views sensitivity to both classes of cues as arising from the same processes: extraction and integration. Thus, acoustic cues influence segmentation in the same way as conditional statistics: via culling chunks from speech. Pauses provide units for extraction that do not require pre-bracketing to be stored as a chunk. Similarly, experience with lexical items changes the types of items that are extracted through the modulation of attention (e.g., storing enough extracted trochees yields a bias to cull chunks from speech that contain trochaic lexical stress).

Some mechanistic accounts of word segmentation conceptualize sensitivity to conditional statistical structure as arising from distinct mechanisms from those that produce sensitivity to acoustic cues such as lexical stress and phonotactic rules (e.g., Mersad & Nazzi, 2011). For example, an account put forth by Shukla et al. (2007) suggested that computations are performed separately on transitional probability information and prosodic information. It is only at a later stage of processing that the output of these computations interacts. Specifically, items characterized by high transitional probabilities are stored as potential word forms. Then, prosodic information is used as a filter to eliminate items that are unlikely to be words on the basis of their prosodic well-formedness (e.g., a cluster of syllables containing a perceptible pause in the middle). Thus, in this model, learners are proposed to perform two separate sets of computations in a hierarchical sequence. In contrast, the Extraction and Integration Framework conceptualizes sensitivity to both classes of cues as arising from the same process. According to a chunking perspective, for an item to be stored as a unified percept its component elements must be held simultaneously in attention, which fits with research that indicates that secondary tasks that demand attention disrupt statistical learning (e.g., Toro et al., 2005; but see Musz, Weber, & Thompson-Schill, 2015) and that attention can modulate the statistical relations that are acquired (e.g., Baker et al., 2004). Thus, even if high transitional probabilities between syllables exist, the presence of an intervening perceptible pause will prevent these elements from being stored in a single chunk. In this way, a chunking account of word segmentation can account for a variety of language-universal cues (e.g., conditional probability, utterance edges, perceptible pauses) invoking only a few simple memory-based processes. These universal cues allow infants to identify a proto-lexicon. This proto-lexicon allows infants to identify the phonological regularities that characterize word forms (via the process of integration). In turn, discovering these phonological patterns changes the process of extraction such that units that are consistent with the pattern are more likely to be segmented.

These language-specific phonological cues eventually supplant statistical structure in importance, as many of them may be more salient to the infant than conditional probabilities (e.g., lexical stress; Jusczyk, Cutler, & Redanz, 1993; Jusczyk & Thompson, 1978; Thiessen & Saffran, 2003, 2007). For example, one study revealed that 8-month-old infants prioritize stress and co-articulatory cues over statistical information about syllable co-occurrence, when the cues were placed in conflict (Johnson & Jusczyk, 2001). Similarly, a second study with adult learners found that incongruent phonotactic cues disrupted the ability to segment based on statistical structure (Finn & Hudson Kam, 2008). Finally, results from a study that pitted phonotactic cues against stress cues found that for 9-month-old infants, stress cues were stronger than phonotactic cues (Mattys, Jusczyk, Luce, & Morgan, 1999). Indeed, a wide variety of experiments have placed statistical cues in conflict with acoustic cues, and most of these experiments demonstrate that by the age at which infants are able to use these cues, they are weighted more heavily than information about syllable co-occurrence (e.g., Finn & Hudson Kam, 2008; Johnson & Seidl, 2009). Note that this does not invalidate the claim that syllable co-occurrence is one of the earliest cues to word segmentation (e.g., Thiessen & Erickson, 2013a). Instead, we suggest that these results are consistent with the Extraction and Integration Framework due to the way in which syllable co-occurrence and acoustic cues interact when processing speech. Once an infant has learned a phonological pattern that is relevant for word segmentation such as lexical stress, these phonological cues can be recruited very quickly in online speech processing (e.g., an infant needs only hear a syllable once to determine whether it is stressed, thus acquiring information relevant for word segmentation). In contrast, the use of statistical cues to word boundaries may require many repetitions before the item becomes robustly represented in memory.

Furthermore, research about how cues conflict fails to address an important aspect of real language processing: no single cue to word segmentation is perfect, and thus segmentation relies on taking advantage of multiple congruent cues to segmentation. Critically, to the extent to which infants are able to take advantage of multiple imperfect sources of information, the available cues can all work together to aid infants in the task of segmenting speech. Indeed, several computational models are consistent with the idea that converging sources of information yield better segmentation performance (Christiansen, Allen, & Seidenberg, 1998; Mattys, White, & Melhorn, 2005; Perruchet & Tillmann, 2010). Studies that focus solely on cue conflict, in the interest of identifying a hierarchy of cue strength, fail to provide a characterization of how sensitivity to these cues develops, and how the use of these cues interact with each other in typical language situations when multiple cues converge rather than artificially conflict.

Although little research has explored how infants might integrate multiple cues, there have been some exceptions to this trend (e.g., Lew-Williams, Pelucchi, & Saffran, 2011; Morgan & Saffran, 1995). For example, Morgan and Saffran (1995) tested 6- and 9-month-old infants' abilities to integrate sequential and rhythmic information while developing word-like representations. In this study, infants listened to syllable pairs separated by silence and a noise detection task was used to determine how cohesive they perceived the syllable pairs to be. The experimenters manipulated whether the sequential and rhythmic information was correlated or uncorrelated. For example, in the correlated condition, the syllable *ga* might always follow the syllable *ko*, and in addition *ga* might always have a longer duration than *ko*. They found that 9-month-olds represented the syllable pairs as a coherent unit only when both rhythmic and syllable-order information supported such a grouping and not when only one of the two cues did. In contrast, 6-month-olds showed evidence of representing the syllable pairs as long as the rhythmic cue supported the grouping, regardless of whether the syllable-order regularity was present. The author interpreted these results as evidence that 9-month-olds are able to integrate both types of information whereas 6-month-olds are not. To perceive a cluster of syllables as coherent, 6-month-olds attended solely to the rhythmic cue. Rhythm has been found to be salient to infants (e.g., at birth infants already show a preference for their native language over languages from a different rhythmic class; Nazzi, Bertoncini, & Mehler, 1998). Thus, salience might explain why rhythm contributed more to 6-month-olds' perception of coherence (e.g., Mehler et al., 1988). By 9 months of age, however, infants have increased abilities to integrate multiple sources of information, which is likely to aid them in more efficiently and accurately segmentation the fluent speech stream, which is rich with imperfect cues to word boundaries. Consistent with this finding, Jusczyk et al. (1999) found that 10.5-month-olds were able to integrate statistical and lexical stress cues to correctly segment the phrase *guiTAR#is* into an iamb followed by an unstressed monosyllabic word, whereas 7.5-month-old infants' overreliance on lexical stress led them to segment the trochaic non-word *TARis* from speech.

A recent study tested infants' ability to integrate statistical cues to word boundary with another source of information: words presented in isolation (Lew-Williams et al., 2011). According to one estimate, 9% of caregiver utterances consist of isolated words (Brent & Siskind, 2001). These isolated occurrences are unlikely to be sufficient to build the infants' early lexicon for several reasons. One reason is that these occurrences, although compelling, are relatively infrequent. Moreover, even words in isolations involve some level of ambiguity, because infants have no way of distinguishing bi- or tri-syllabic utterances that are words (e.g., *tiger*; *happiness*; *finishing*; *Saturday*) from utterances that are composed of multiple mono- and bi-syllabic words (e.g., *I love cake*; *I'm tired*; *You're feverish*; *I'm coming*) without some additional source of knowledge. One possible source of information is statistical co-occurrence information. Brent and Siskind (2001) found that of the 9% of isolated words in the input, 27% of words appear two or more times in neighboring sentences (e.g., *See the doggie there? Doggie!*). Similarly, Onnis, Waterfall, and Edelman (2008) found that 58.6% of sentences in child-directed speech share lexically redundant information with neighboring sentences.

If infants were to integrate statistical co-occurrence information with pause information, this would likely prove a viable strategy for the efficient segmentation of speech. According to the Extraction and Integration Framework, pauses serve as a bracketing aid for the chunking process, because sound combinations that are separated by pauses will be less likely to be held in attention to be chunked. In this way, the framework accounts for integration of pause information with other cues using the same processes (i.e., chunking adjacent sound sequences held jointly in attention) rather than using a set of separate processes. Lew-Williams et al. (2011) investigated the speech segmentation abilities of the 9.5-month-old infants using natural language stimuli. Critically, they increased the difficulty of the task by shortening the exposure to the nonsense language to prevent infants from successfully segmenting the speech on the basis of statistical structure alone. In the *Fluent Speech Only* condition, infants heard only fluent speech containing 12 repetitions of each target word. In a *Fluent Speech + Isolated Words* condition, infants heard fluent speech containing 6 repetitions of each target word. In addition, they heard three isolated tokens of each word. Subsequently, they tested infants' ability to discriminate between high-probability items and low probability items. They found that infants in the *Fluent Speech + Isolated Words* condition discriminated high- from low-probability words, whereas infants in the *Fluent Speech Only* condition failed to discriminate. Critically, the high- and low-probability items were heard an equal number of times in isolation. Thus, the superior recognition of high-probability items found in the *Fluent Speech + Isolated Words* condition must have reflected an integration of the two sources of information. Items segmented statistically, then, can be integrated with words in isolation. By the model of word segmentation advanced here (Thiessen et al., 2013), words in isolation can be conceptualized as chunks stored in memory after they were presented bracketed by perceptible pauses.

Thus, a growing body of evidence suggests that sometime after the first half of the first year of life, infants begin to be able to integrate multiple sources of information in the context of word segmentation. This ability allows infants to benefit from multiple imperfect sources of information to converge upon accurate parsings of speech. Here, we have described this process in terms of extracting discrete and coherent chunks from the input and integrating across those chunks to learn phonological patterns that drive later segmentation (e.g., Giroux & Rey, 2009; Thiessen & Erickson, 2013b; Thiessen et al., 2013). According to chunking accounts of word segmentation, some global acoustic cues will also emerge early and influence what units are extracted, as a result of the nature of human memory and information processing. In particular, attention will be the driving factor, which is consistent with evidence that statistical learning cannot proceed in the absence of attention (Baker et al., 2004; Toro et al., 2005). Thus, because clusters of sounds must be held simultaneously in attention to be grouped, sounds which cross a phrasal boundary or contain a perceptible pause will not be grouped (see Baker et al., 2004 for evidence of this principle in the context of visual statistical learning). In contrast, infants will not be able to integrate other language-specific acoustic cues (e.g., phonotactics) with other information until they have acquired that regularity via an analysis of the central tendency of stored word forms. Once acquired, language-specific knowledge of prosodic patterns can serve to bias attention, such that units that are consistent with that knowledge tend to be extracted (e.g., learners will be more likely to extract sound clusters that obey phonotactic rules than clusters that violate those rules).

Beyond word segmentation

Since the discovery of statistical segmentation, many studies have replicated infant segmentation on the basis of statistical structure (e.g., Johnson & Jusczyk, 2001). Other studies have extended these findings to other participant groups: adults (e.g., Saffran, Newport et al., 1996), children (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997), and neonates (Bulf, Johnson, & Valenza, 2011; Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009). Thus, statistical learning is intact very early in infancy; this is critical to predictions made by statistical learning accounts, which argue that sensitivity to the statistics of the input bootstraps the acquisition of language-specific regularities. Furthermore, the representations formed from statistical learning are long lasting, further supporting predictions that learning of statistical patterns could support long-term learning tasks such as language acquisition (Arciuli & Simpson, 2012a; Kim, Seitz, Feenstra, & Shams, 2009). Perhaps most importantly, as we have argued above, statistical learning applies to more kinds of statistical structure than simply the sequential co-occurrence probabilities between syllables, and can be observed with both other kinds of linguistic regularities and many other kinds of structured non-linguistic input (e.g., Conway & Christiansen, 2005; Fiser & Aslin, 2001) and over many kinds of structured input (e.g., Baldwin, Andersson, Saffran, & Meyer, 2008; Creel et al., 2004; Fiser & Aslin, 2001; Kirkham, Slemmer, & Johnson, 2002).

The flexibility and generality of statistical learning suggests that it may be useful for other aspects of language learning beyond word segmentation. One goal of the Extraction and Integration account is to extend statistical learning accounts of language acquisition beyond word segmentation, in a unified framework. As an example, consider demonstrations that statistical learning applies not only to the acquisition of adjacent regularities (e.g., one syllable predicted the syllable directly following) but also to the learning of nonadjacent regularities – regularities that exist over an intervening element (e.g., Newport & Aslin, 2004; Thompson & Newport, 2007). The acquisition of syntactic structure also relies on sensitivity to long-distance dependencies (e.g., the keys to the cabinet are on the table). Sensitivity to nonadjacent statistical regularities has been proposed as a mechanism underlying syntactic development (e.g., Gomez & Gerken, 1999; Newport & Aslin, 2004). Both adult and infant learners are able to acquire regularities at the level of word order (Kaschak & Saffran, 2006; Saffran, 2001b; Saffran & Wilson, 2003). From our perspective, syntax acquisition based on nonadjacent regularities arises from the same fundamental processes as the discovery of words and phonological patterns from adjacent statistical regularities: extraction and integration. One primary difference between discovering words and syntax according to this account is that high variability is needed in the elements that intervene between nonadjacent dependencies (e.g., Gómez, 2002). As a result, integration is recruited in the discovery of syntactic relations to a greater extent than in the discovery of words (Thiessen & Pavlik, 2013).

This perspective contrasts with traditional linguistic approaches to the distinction between syntax and the lexicon (e.g., Pinker, 1998; Pinker & Ullman, 2002). Instead, this perspective shares much in common with connectionist and usage-based approaches (e.g., Elman, 1998; Goldberg, 2003; Tomasello, 2000; Tomasello, 2003) and argues that language is composed of statistics at many different levels. Learners adapt to statistics at these levels as a function of the same underlying processes that are involved in word segmentation. According to this approach, phrase-level regularities are discovered in much the same way as phonological regularities are discovered from the extraction of words. Learners extract multi-word chunks (e.g., *they're buying it*, *they're getting it*, *they're hiding it*), and then integrate across those stored chunks to discover a phrase level regularity (e.g., *they're VERBING it*). Rather than composing syntactic constructions based on stored rules, initial schemata are lexically specific; groups of words are clustered into groups based on phonological and semantic similarity. The strength of a particular cluster of words is a function of both the size (i.e., the number of words in the cluster) and similarity between the cluster and new lexical item to be assimilated into the lexical schema. Eventually, when enough items have been assimilated into these lexical frames, they appear abstract and rule-like.

Existing empirical evidence is consistent with statistical learning or usage-based approaches to syntax learning. Usage-based accounts predict that initial frames should be lexically specific and should only gradually begin to appear abstract. This is consistent with evidence that although 2-year-olds will only reproduce ungrammatical word orders in 34% of their utterances with a familiar verb, they are willing to produce ungrammatical word orders in as much as 69% of their utterances with novel verbs, which suggests their knowledge of grammatical structure is grounded in specific lexical items (Abbot-Smith, Lieven, & Tomasello, 2001). Also relevant is a study that reported that children demonstrate familiarity effects for more frequent multi-word phrases, even when overall item frequency is equated (Bannard & Matthews, 2008). These findings, along with anecdotal reports that infants often undersegment utterances, treating high frequency multi-word phrases as single units (e.g., *more milk, what's this*; Brown, 1973; Reich, 1986), fit with a story in which infants and young children are storing and integrating across multi-word chunks to acquire phrasal structure. Finally, Onnis and Thiessen (2013) reported that native-language phrase level regularities influence both English- and Korean-speaking learners' word segmentation, a finding which might suggest that the same processes are involved in discovering regularities at both word and phrasal levels.

Two other extensions of the Extraction and Integration account to language acquisition phenomena beyond word segmentation are worth noting. One possible extension of the Extraction and Integration Framework is to provide an explanation of findings from the cross-situational word learning paradigm, in which labels are paired with multiple referents and learners can only appropriately map a label to its referent by tracking the co-occurrence probabilities over time (e.g., Yurovsky, Yu, & Smith, 2012; Yu & Smith, 2007). Early in learning, children do not know how words and referents pair together, and extract word–object chunks at random (or as guided by social attention; e.g., Pereira, Smith, & Yu, 2014). Many of these associations will be spurious. These extracted chunks interact with long-term memory representations stored in the cortex – a process we expect to be hippocampally mediated (e.g., McClelland et al., 1995) – such that statistically coherent chunks are strengthened by the process of integration. Over time, word–object representations that are stable, abstract enough to be generalized (because features associated with specific instances are likely to be lost from the overall representation; Thiessen & Pavlik, 2013), and statistically coherent such that they represent pairings between words and objects that are likely to co-occur.

Another extension of the Extraction and Integration account concerns phoneme learning. At birth, infants discriminate beyond many phonetic distinctions that are not phonemic in their native language; by the end of the first year of life, they primarily distinguish between contrasts that are phonemic (i.e., meaningful) in their native language (Werker & Tees, 1984). One possible explanation for this developmental phenomenon is that infants are sensitive to the statistical distribution of phonetic exemplars in the input (e.g., Maye et al., 2002). The phonemic categories of the native language shape the distribution of phonetic exemplars in the input because speakers are less likely to produce phonetic tokens that are ambiguous (that is, tokens that fall between two phonemic categories), and more likely to produce tokens that are good exemplars of the phonemic distinctions in the language, especially in infant-directed speech (Werker et al., 2007). This leads to a peaked distribution in which tokens near the prototypical center of a phonemic category outnumber tokens farther away from the center (i.e., ambiguous tokens). If these exemplars were extracted from the input and stored, subsequent integration over these tokens will yield representations of phonetic tokens in which prototypical clusters of phonemes are more strongly represented than ambiguous tokens (Thiessen & Pavlik, 2013). As such, the processes of extraction and integration may help infants identify the phonemic categories of their native language. In turn, these emerging phonemic representations may constrain subsequent extraction, leading infants to be better suited to learning words in their native language (cf. Emberson, Liu, & Zevin, 2013).

Comparison to other models of statistical learning

In the section that follows, the Extraction and Integration will be compared to other models of statistical learning. Although a comprehensive review of models of statistical learning is outside the scope of this review, the Extraction and Integration Framework will be compared to a selection of models with the goal of highlighting some of the similarities and differences between the framework and other accounts. The clearest difference between the Extraction and Integration Framework and other accounts is that the majority of other models attempt to explain only sensitivity to conditional structure (but see Adriaans & Kager, 2010 for an exception that also explains sensitivity phonotactic patterns through a generalization parameter that shares some similarities with the process of integration, although it accomplishes this generalization using very different processes based on optimality theory; e.g., Prince & Smolensky, 1997). In contrast, the Extraction and Integration Framework explains sensitivity to condition structure via the process of extraction, and sensitivity to distributional properties via the process of integration. However, given that these two processes can be modeled using two separate computational models, I will focus my comparisons on the Extraction and Integration Frameworks' explanation of sensitivity to conditional structure (based on extraction) to other accounts of conditional statistical learning.

Boundary-finding vs. clustering models

Most computational models of sensitivity to conditional information can be classified as either boundary-finding or clustering models. The Extraction and Integration Framework invokes a chunking mechanism to explain the learning of conditional relations, which can be characterized as a clustering model. Clustering models assume that learners are extracting and storing discrete representations. This is accomplished via storing clusters of statistically related elements (e.g., elements with strong conditional relations) in a lexicon of potential word forms (see Orbán, Fiser, Aslin, & Lengyel, 2008 for an example of a clustering model intended to explain visual conditional statistical learning). Clustering models vary in the mechanisms they invoke to explain the storage of units (e.g., chunking; Perruchet & Vinter, 1998; Bayesian hypothesis testing; Frank, Goldwater, Griffiths, & Tenenbaum, 2010), but they share the assumption that discrete representations are extracted and stored. Thus, they explain the discrimination showed by learners of words from artificial languages from foils with lower conditional probabilities by the fact that the words but not statistically incoherent foils are posited to be stored in a lexicon. One advantage of this assumption is that it provides a natural fit to the concept of learning word forms.

In contrast, boundary-finding models operate by searching for regions in the input where the conditional relations between nearby elements are low. Boundaries are inserted at these points of low probability transitions. As an example, some simple recurrent networks are trained to predict elements in a sequence of input (e.g., speech sounds) on the basis of the previous element (e.g., Cairns, Shillcock, Chater, & Levy, 1997; Christiansen et al., 1998; Elman, 1990). When the predictability of an upcoming element is lower than some preset threshold, or when an error signal is high because the predictability in a certain area is poor, a word boundary is inserted. This strategy contrasts from clustering models in that rather than storing units, they represent the patterns of statistical relations between elements in an input. That is, although boundary-finding models vary in the kind of statistical information they use to posit boundaries (e.g., Bayesian statistics; mutual information; e.g., see Frank et al., 2010; Swingley, 2005 for more information), what they share is the assumption that the process of learning is fundamentally about discovering where the input is predictable and where it is not, which is then used to insert a boundary, rather than forming discrete representations. For example, the StaGe model accounts for sensitivity to conditional probabilities via a boundary-finding strategy that relies on observed and expected probabilities (Adriaans & Kager, 2010). Many boundary finding models have been instantiated in connectionist networks (e.g., Simple Recurrent Networks; Christiansen et al., 1998), although connectionist models of conditional sensitivity are not necessarily boundary-finding models. Rather, some connectionist models of conditional sensitivity are better characterized as clustering models because they invoke processes that result in discrete units (e.g., Boucher & Dienes, 2003; French, Addyman, & Mareschal, 2011).

Clustering models are able to explain the finding that learners respond differentially to units and subcomponents of units (e.g., knowing the word *helicopter* makes it difficult for learners to recognize the subcomponent *heli*; Giroux & Rey, 2009), because as learners become more familiar with a unit, the subcomponents embedded in that unit are removed from the lexicon. This is accomplished by competition between the stored units, although different models achieve this competition using different processes (e.g., Frank et al., 2010; Orbán et al., 2008; Perruchet & Vinter, 1998). In contrast, boundary-finding models do not invoke processes that result in competition between units and their embedded components, and are consequently unable to account for the finding that learners become worse at identifying embedded components of units as they become familiar with the whole unit. That is, because embedded units have equally high conditional relations as the whole unit, they predict that learners should be equally good at identifying the embedded unit.

An important consideration relates to the kinds of computations that models of statistical learning perform to account for learning of statistical structure. Some models explicitly compute transitional probabilities (e.g., Gambell & Yang, 2004; Mersad & Nazzi, 2011), based on the typical assumption that learners compute transitional probabilities between elements (Saffran, Aslin et al., 1996). In contrast, other models do not explicitly compute transitional probabilities, despite the fact that they invoke processes that lead to the appearance of sensitivity to transitional probability (e.g., Perruchet & Vinter, 1998; Servan-Schreiber & Anderson, 1990). For example, PARSER can learn structure that can be characterized by transitional probabilities between elements, despite the fact that it merely stores frequently appearing sequences as chunks. An advantage of the latter kind of approach is that it shows greater psychological plausibility approaches that explicitly calculate transitional probabilities. Transitional probability approaches are also particularly vulnerable to concerns regarding the units of computation. This is because if learners calculate transitional probabilities between elements, it becomes necessary to specify between which elements learners calculate transitional probabilities (e.g., phonemes or syllables). Although some researchers have argued that syllables are likely to be the unit for computation (e.g., Bertoncini & Mehler, 1981), based on studies that report that infants find syllables to be more salient than phonemes (e.g., Bijeljac-Babic, Bertoncini, & Mehler, 1993), other approaches have argued that neither phonemes nor syllables are psychologically real (e.g., Goldinger & Azuma, 2003). Chunking approaches are less vulnerable to this concern because a chunk can be described as a sequence of syllables or phonemes or an acoustic episode of some duration.

Computational architectures

Computational models of statistical learning can be implemented in a variety of architectures, such as Bayesian, connectionist, or symbolic/hybrid symbolic models. The architecture in which models of statistical learning are implemented is independent of the clustering/boundary finding distinction. For example, although many boundary-finding models are implemented in connectionist architectures, connectionist models can also implement clustering processes (e.g., French et al., 2011).

Although architectures are largely independent of the strategy that is used to achieve sensitivity to statistical structure, different architectures are often differentially suited to addressing different kinds of questions. For example, an advantage of connectionist architectures is that they possess a certain amount of biological plausibility, given that learning involves patterns of activity among many low-level interconnected units that exhibit some similarity to neurons. Connectionist architectures can be thought of as describing an algorithmic level of learning according to Marr's levels, in which the actual processes involved in learning are specified (Marr & Poggio, 1979). In contrast, other models may describe learning at a computational level, in which the goal is not to describe the specific algorithms used but rather the functions that those algorithms should attempt to explain. For example, normative descriptive statistical models do just this. Shortlist B is a Bayesian model of recognition of continuous speech that can be described as a model at the computational level rather than at the algorithmic level (Norris & McQueen, 2008). Similarly, the DR model described by Brent and Cartwright (1996) achieves segmentation based on evaluating probabilities in a batch process that relies on having memorized a large segment of input (note that this is not a necessary feature of Bayesian models; rather, some Bayesian models may be characterized as process models, and implement constraints that may be akin to memory limitations (e.g., INCDROP; Cartwright & Brent, 1997; Frank et al., 2010).

Chunking models are a type of clustering model that have generally been implemented in symbolic or symbolic/hybrid frameworks (e.g., PARSER; Perruchet & Vinter, 1998; Competitive Chunker; Servan-Schreiber & Anderson, 1990). However, although Servan-Schreiber and Anderson's (1990) Competitive Chunker model was originally formulated as a symbolic model, it has also been implemented in a connectionist network (Boucher & Dienes, 2003). Although chunking processes can be simulated using non-symbolic processes, a common feature to all chunking models is that they are necessarily clustering models rather than boundary-finding models, given that they result in the storage of discrete representations. Historically, chunking models have been more commonly applied to the adult Artificial Grammar Learning literature (e.g., Reber, 1967), whereas perspectives that argue for statistical computations between elements have been prominent in the infant statistical learning literature (e.g., Saffran & Wilson, 2003; Perruchet & Pacton, 2006). The Extraction and Integration Framework constitutes a departure from this traditional distinction, given that it originated from an infant statistical learning perspective, but it invokes chunking processes to explain sensitivity to statistical structure.

Breadth of phenomena

A final distinction between the Extraction and Integration Framework and other models of statistical learning lies in the breadth of phenomena that they attempt to explain. Whereas the previous sections have been largely concerned with differentiating between the processes and architectures of various models of conditional statistical learning, the Extraction and Integration Framework is unique in that it attempts to account for the learning of many statistical regularities beyond conditional structure. One exception to this is the StaGe model, which does include a generalization parameter that allows incorporating information such as phonotactics into the learning processes (Adriaans & Kager, 2010, although it achieves this using explicitly linguistic processes rather than the domain-general process invoked by the Extraction and Integration Framework). Whereas almost all of the models described previously have been applied exclusively to conditional probability learning, the Extraction and Integration Framework attempts to describe category learning (e.g., learning phonemes based on the frequency and variability of a set of exemplars; Maye et al., 2002) as well as the learning of cue-based statistics (e.g., learning to segment based on stress; Thiessen & Erickson, 2013a; Thiessen & Saffran, 2007). Thus, the greatest advantage of the Extraction and Integration Framework may not lie in the particular algorithms that it favors as a description of the processes involved in the acquisition of statistical structure, but rather the idea that statistics other than conditional probability exist and must be explained.

In addition, although models of statistical learning may have generally focused exclusively on the learning of conditional relations, the processes relevant for the acquisition of distributional statistical structure have been studied in the context of the category learning literature. Other models of category learning may be equally good at explaining distributional statistical learning. For example, the Extraction and Integration Framework favors iMINERVA, an exemplar memory model that shares some similarity with prototype models insofar as it generates aggregate representations that are interpretations of sets of exemplars or experiences. However, although exemplars are preferred because they make an immediate point of contact with the idea of storing word forms, this preference does not represent a strong theoretical commitment; iMINERVA might be implemented as a model that stores aggregrate representations rather than specific exemplars. Potentially relevant is the research from the probabilistic category learning literature, which has largely focused on learning with explicit feedback that taps into the basal ganglia systems (e.g., Maddox & Ashby, 2004). It is unclear whether the neural systems recruited in these paradigms are distinct from those recruited in statistical learning tasks, in which no feedback is given. However, although no explicit feedback is given in traditional statistical learning tasks, the Serial Reaction Time Task (Nissen & Bullemer, 1987) may be conceptualized as involving a source of incidental feedback given that the participants show a decrease in the amount of time and the number of errors in pressing the keys as they learn the pattern. Research indicates that the basal ganglia are recruited in some tasks that involve incidental rather than explicit sources of feedback (Lim, Holt, & Fiez, 2013). In addition, some neuroimaging studies have reported that basal ganglia activation is seen during some statistical learning tasks (Karuza et al., 2013; McNealy, Mazziotta, & Dapretto, 2006). It remains to be seen whether and how these learning systems are related to the learning involved in statistical learning paradigms in which the learner receives no explicit feedback.

Evaluating models of statistical learning and their predictions

There are many models of statistical learning, and these models achieve sensitivity to statistical structure using a wide variety of processes (e.g., Perruchet & Vinter, 1998; Frank et al., 2010; French et al., 2011). The processes that these models invoke may be useful in deriving criteria to compare the benefits and drawbacks of various models. One such criterion is psychological plausibility, based on what is known about the cognitive architecture in which these models ultimately must operate. One advantage of a chunking model such as PARSER (Perruchet & Vinter, 1998) is that it can explain sensitivity to conditional structure without explicitly calculating transitional probabilities as some models do (Mersad & Nazzi, 2011). PARSER simply pulls out chunks from the input and can learn the structure of input that can be described using transitional probabilities without the model ever learning anything about transitional probabilities. This also constitutes an advantage relative to Bayesian perspectives that use hypothesis testing processes, because even with memory constraints imposed, hypothesis testing is somewhat implausible as an explanation of infant behavior even if it may provide a reasonable approximation of adult performance. Another advantage of chunking approaches such as PARSER is that they have the advantage of making contact with the idea of word forms by positing the formation of a discrete representation, whereas boundary finding approaches search for boundaries, which appears less compatible with the idea of learning words at least on a surface level. Finally, chunking models such as PARSER provide a better fit to the data reported by Giroux and Rey (2009) than boundary-finding models (e.g., an SRN model; Elman, 1990), which demonstrated that as learners become more familiar with chunks, they become less familiar with the sublexical components (e.g., becoming less familiar with *eleph* as *elephant* is better learned; see also Slone & Johnson, 2015, and Orbán et al., 2008 for evidence of the same principle in visual statistical learning). Thus, the principles that different models use to achieve sensitivity to statistical structure may be more informative with respect to which model provides the best fit to human performance than predictions regarding connections between learning of statistical relations and acquisition of real language.

Some models of statistical learning make different predictions from each other by virtue of the different processes they invoke to account for the learning of statistical structure. Specifically, models differ in the extent to which they predict that statistical learning should be related to or independent of the aspects of domain general cognition. For example, because StaGe generalizes by invoking principles from optimality theory, it uses explicitly linguistic processes (Adriaans & Kager, 2010; Prince & Smolensky, 1997). For models that use explicitly linguistic processes, Thiessen has argued that generalization in language should follow different patterns than generalization with nonlinguistic stimuli (Thiessen, 2011; Thiessen et al., 2013). Contrary to this prediction, Thiessen has argued that linguistic generalization at least in some forms occurs because of domain-general processes of similarity rather than domain-specific constraints (Thiessen, 2011). In addition to making arguments about domain-generality, PARSER and the Extraction and Integration Framework both posit that sensitivity to conditional structure arises from general attention and working memory processes. This stands in stark contrast to perspectives that argue that statistical learning is independent of the processes such as attention and working memory (e.g., traditional implicit learning perspectives; e.g., Hayes & Broadbent, 1988). For example, Ullman's Declarative–Procedural Model makes a clear distinction between lexical and syntactic processes (e.g., Ullman, 2004). Ullman (2004, 2005) has argued that whereas facts are learned via declarative and working memory systems, the procedural memory system underlies the learning of statistical structure, which is independent of declarative and working memory processes and structures (e.g., Lum, Conti-Ramsden, Page, & Ullman, 2012). Similarly, although TRACX is a connectionist model that bears more resemblance to chunking models such as PARSER than many of the other connectionist models that implement boundary-finding strategies, it also predicts the independence of statistical learning from working memory (French et al., 2011; French & Cottrell, 2014).

Summary

Numerous models have been proposed to explain the process underlying statistical learning (e.g., Christiansen et al., 1998; Shukla et al., 2007). The goal of this section was to provide a description of the Extraction and Integration Framework and discuss commonalities with and differences from other models of statistical learning, both with respect to the processes that they invoke and the kinds of predictions that they make. Whereas many models of statistical learning are sensitive only to conditional information (e.g., Cairns et al., 1997; Christiansen et al., 1998) and may frequently explicitly calculate transitional probabilities (e.g., Adriaans & Kager, 2010; Shukla et al., 2007), the main contribution of the Extraction and Integration Framework is the suggestion that statistical learning consists of two major processes that together explain how learners acquire many aspects of statistical structure. Extraction fundamentally involves a chunking process in which frequently occurring sequences are likely to be chunked into discrete units. Integration involves similarity-weighted aggregation over stored chunks to induce some aspect of central tendency. Critically, this learning then biases the extraction parameter, such that learning influences the kind of chunks that are likely to be subsequently extracted. One main advantage of this conceptualization of statistical learning is that it can explain more than just sensitivity to conditional probabilities.

Here, we have described a mechanistic account of statistical learning and provided an explanation of how sensitivity to statistical cues might be integrated with other cues in the context of language acquisition. In addition, we have provided a brief discussion of how this model compares to other models of statistical learning, which have largely attempted to explain sensitivity to conditional probabilities. The theoretical account that we have described goes beyond the conceptions of statistical learning as a transitional probability calculator. Instead, statistical learning is characterized as a mechanism responsible for the learning of a much broader class of phenomena. We suggest that this mechanism is, in fact, an important driver of early language acquisition. However, making such a claim requires more than a simple demonstration and definition of a learning mechanism. Additionally, it requires evidence that statistical learning is linked to "real" language acquisition. Throughout the rest of this review, we will attempt to provide such evidence. We will begin by addressing one of the most powerful possible objections to our theoretical stance: namely, that the kinds of stimuli used in laboratory demonstrations of statistical learning are too artificial and too distant from real linguistic input to be informative with respect to the process of language acquisition.

Laboratory studies of statistical word segmentation

Given that laboratory studies of statistical word segmentation suggest that statistical learning is a rapid domain-general learning mechanism, researchers have speculated about the role this mechanism might play in the acquisition of natural languages (e.g., Mirman, Graf Estes, & Magnuson, 2010; Newport & Aslin, 2000). With respect to word segmentation, although there are numerous studies that demonstrate infant sensitivity to statistical cues to word boundaries (e.g., Graf Estes, Evans, Alibali, & Saffran, 2007; Graf Estes, Evans, & Else-Quest, 2007; Saffran, Aslin et al., 1996), they have necessarily involved an extremely simplified input because stimuli must contain no other cues to word identity to isolate the effect of statistical structure on word segmentation. In contrast, naturally produced speech contains many such cues, which must be removed to observe the effect of any particular cue on segmentation; many of the early studies of word segmentation did just this. As a result, some researchers have speculated that learning that is commonly found in laboratory settings with stripped down artificial stimuli is unlikely to be related to processes involved in the acquisition of natural languages: although the studies may themselves be soundly designed, the learning involved is intimately tied to experimental input. According to this view, the learning process observed in these studies is insufficiently robust to scale up to the complexity found in natural languages (e.g., Johnson & Seidl, 2009; Johnson & Tyler, 2010).

Ecological limitations of early studies of statistical learning

Early experimental studies of statistical learning were limited in their ecological validity, which has generated concerns that performance on statistical learning tasks does not relate to real language processes (e.g., Endress & Mehler, 2009; Johnson & Tyler, 2010). In this section, we will focus on the statistical word segmentation literature for two reasons. First, investigations of statistical learning began in the context of word segmentation and thus most of the early work on statistical learning is focused on word segmentation. Second, the statistical word segmentation paradigm is perhaps the most artificial of the experimental paradigms exploring statistical learning (given the necessary dissimilarity between the vast amounts of word segmentation experience children have in natural language compared to laboratory settings), such that many of the ecological validity criticisms apply uniquely or particularly to word segmentation. These limitations in ecological validity can be conceptualized as falling into two broad categories with respect to the stimuli they used: (1) the stimuli are insufficiently complex and (2) the stimuli are unnatural. To some degree these issues overlap in the context of natural language, where the extent to which an item is complex and naturalistic is fundamentally intertwined; however, these constructs are logically separable and represent distinct concerns about the scalability of findings. To address concerns regarding the ecological validity of statistical learning, both the complexity and the naturalism of the stimuli used in these studies must be addressed. Although the nature of the statistical learning segmentation paradigm is such that certain limitations in ecological validity are difficult if not impossible to overcome, remarkable strides have been made with recent studies to use materials that more closely resemble the infant's real language input (e.g., Pelucchi, Hay, & Saffran, 2009; Saffran & Wilson, 2003; Thiessen et al., 2005).

Complexity

Artificial language stimuli used in early studies of statistical segmentation were simplified relative to real language with respect to (1) the acoustic properties of items used (e.g., whether the utterances followed natural pitch contours) and (2) the distribution of the words in the experimental input (e.g., the particular order of the words in the miniature language). As will be discussed, increasing the complexity of the acoustic properties of the words themselves within the typical statistical learning paradigm is a relatively feasible task. In contrast, the simplified nature of the distribution of words in the experimental input will likely prove to be a greater challenge. Alternative and likely indirect approaches (e.g., correlational methods) may be necessary to address concerns regarding the complexity of the distribution of words.

Early statistical learning experiments used speech stimuli whose acoustic properties were simplified in key ways. Natural speech contains multiple informative cues, both language-general and language-specific. Rhythmic and stress cues, phonotactic regularities, and pauses at word boundaries may all contribute to the identification of word boundaries. Consequently, these regularities were removed from early speech stimuli, the better to demonstrate that infants could segment fluent speech solely on the basis of probabilistic information about syllable co-occurrence. In addition to lacking regularities that may be useful to word boundaries, such speech stimuli also lacked sources of acoustic variability that are present in natural language but that do not inform word identities. Specifically, individual *tokens*, or productions of words, exhibit dramatic acoustic variability. This variability differs depending on many features (e.g., speaker gender and identity). For example, tokens of words in speech vary based on *co-articulation*; phonemes differ depending on the acoustic realization in which they are produced: a /d/ actually sounds acoustically distinct depending on which phoneme precedes and follows it.

Both types of acoustic variation – that which informs word boundaries and that which adds uninformative (with respect to word boundaries) complexity – are present in real language, and their absence in statistical learning experiments raises questions about how the learning in the laboratory maps onto learning in naturalistic settings, albeit in different ways. One type of variability (i.e., that based on indexical characteristics or variability in particular productions of word forms) may simply make the learning process more difficult by requiring that learners recognize word forms in the face of acoustic variability. The other kind of variability, involving additional cues to word boundary, may actually simplify the learning process. However, the inclusion of this informative variability in laboratory stimuli frequently produces results that are agnostic about the relative contributions of particular cues. The earliest studies of statistical learning sought to isolate effects of statistical structure, which meant removing these many potential sources of information. To this end, these studies eschewed natural speech stimuli in favor of synthesized speech stimuli, which were constructed in such a way to remove all extraneous cues. What resulted was a mini-language design that – because it stripped away all other possible cues – allowed researchers to make causal claims about the role of conditional statistics in the segmentation of speech. However, this removal of cues meant that the acoustic properties of the words themselves were simplistic relative to the acoustic richness of words found in natural input. Note that (as we will discuss further in the next section, “naturalism”) subsequent studies have rectified this shortcoming by using more complex acoustic input (e.g., Thiessen et al., 2005).

The frequency and the distribution of the lexical items – like the acoustic properties of the speech itself – have also been simplified in traditional studies of statistical word segmentation. In these studies, conditional probabilities must be manipulated to conclude that statistical structure has a causal role in speech segmenting. Thus, artificial languages are created where conditional probabilities of syllables are structured to create high-probability items (i.e., words; syllables that co-occurred regularly) and low-probability items (e.g., part–word foils; syllables that co-occurred only incidentally between two words). Because practical time limitations preclude exposure phases that are inordinately long when testing infants, statistical regularities in experimental language input must be compressed into mini-languages, else experiments would persist for hours or days. The length of the exposure varies depending on the age group of the participants. Whereas with adults and children, exposure lengths can be 15 minutes or longer (e.g., Saffran et al., 1997), with infants, the length of exposure is necessarily brief (typically 1–2 minutes; Saffran, Aslin et al., 1996; Thiessen et al., 2005) due to their limited attentional capacities (e.g., McCall & Kagan, 1970).

Despite this brevity, evidence of learning has been demonstrated in young infants after only 2 minutes of exposure to mini-languages (e.g., Saffran, Aslin et al., 1996). The consequences of this time constraint on the design is twofold: the input is a highly concentrated presentation of only a few words, and the conditional predictability of syllables that comprise words is perfect because no words are permitted to share component syllables. This is typically achieved through the presentation of 4–6 words (groups of statistically coherent di- or tri-syllables) with the constraint that no word can repeat in immediate succession. In this exposure, the transitional probability between two syllables in a word is 1.0, and the transitional probability between two syllables that span a word boundary is 0.33 or less (e.g., in a language with four words, if *bidaku* and *golabu* are words, the transitional probability between *bi* and *da* is 1.0, whereas the transitional probability between *ku* and *go* is 0.33). Such idealized conditional probabilities contrast with real language input, in which infants experience utterances containing many distinct words, with fewer repetitions close in temporal proximity. Similarly, because real words contain overlap in the particular syllables they comprise, the conditional probabilities that inform word boundaries are necessarily noisier. However, the presence of this noise is likely mitigated by the fact that the amount of speech that infants hear in real language input is vastly greater than what they hear in simplified artificial languages used in the laboratory.

The presence of noise in the statistical regularities found in real language can be used to derive predictions about the rate of vocabulary development. For example, because the difference in the level of statistical coherence between words and non-words is less sharply marked in real languages than in artificial languages, infants should make some errors. They should fail to recognize some words with low statistical coherence and should treat some non-words with high statistical coherence (e.g., phrases like *good morning*) as though they are real words. Evidence suggests this is the case at least for young infants learning French (Ngon et al., 2013). Similarly, the noisy statistical structure of real languages means that word learning should exhibit a relatively protracted learning trajectory (e.g., Yurovsky, Fricker, Yu, & Smith, 2013). In other words, the noise in the input necessitates that infants must hear many utterances before the higher probabilities between syllables that compose words become evident.[3] Once infants have identified a set of words, the presence of these recognized forms in the input, and any learned phonological regularities, will simplify the learning problem and make learning proceed more rapidly.

Computational models may provide insight into this issue. A computational model of vocabulary development that relied on sensitivity to statistical regularities in label-referent mappings successfully simulated key features of naturalistic vocabulary growth trajectories; although the model was endowed with the capacity to detect statistical regularities immediately (i.e., from birth), it did not begin to learn words until relatively late, and the rate of learning increased over time (Yu, 2008). This learning trajectory accords with empirical evidence that suggests that vocabulary development begins slowly and accelerates during the second year of life (Bates, Bretherton, & Snyder, 1988; Gopnik & Meltzoff, 1987; Lifter & Bloom, 1989). Evidence such as the computational model described in Yu (2008) indirectly buttresses the claim that the learning from such statistical mechanisms can scale up to the complexity of natural language, by demonstrating that models instantiating statistical word-learning algorithms show a similar trajectory to real vocabulary development curves.

Concordant with such a claim is evidence from corpus analyses (e.g., Gervain & Guevara, 2012; Perruchet & Vinter, 1998; Swingley, 2005). The application of segmentation algorithms using statistical co-occurrence information to real language corpora demonstrates that the use of this type of computational strategy by learners is feasible; segmentation algorithms sensitive to statistical structure yield successful segmentations of speech. Despite this support for the plausibility of statistical mechanisms in accounts of children's vocabulary development trajectories, few laboratory studies have tested directly whether learners can cope with noisier statistics that more closely approximate features of real language input. A notable exception to this is a study that investigated segmentation with noisier statistics, which found that adults were still able to demonstrate evidence of segmentation (Frank et al., 2010). However, as this study only tested adults, an important task for future studies will be to investigate whether infants are able to discover words when the statistics of the input are imperfect. This question may require a longitudinal design to ensure that the exposure is long enough for infants to acquire the statistical structure among the noise.

Complexity in the form of noisy statistics, as described above, almost certainly poses a challenge to word segmentation based on statistical structure. This is consistent with the general assumption (often implicit) that greater complexity is harmful to infant learning. For example, more complex stimuli require longer amounts of processing time for infants to encode than less complex stimuli, which suggests that complex items may be more difficult to process than simple item (Hunter, Ames, & Koopman, 1983). One aspect of speech directed at infants and young children is that it tends to be simplified and redundant in its structure relative to speech directed at older children and adults (Snow, 1972). Similarly, Newport's (1990) *less-is-more* theory of language development posits that infants are actually more successful at language acquisition than adults because their limited cognitive capacity forces

[3] Factors in addition to the noise in the input, such as fine turning native language phonemic categories, also likely contribute to the protracted developmental trajectory of vocabulary development (e.g., Kuhl, 2004; Werker & Tees, 1984).

them to process a truncated portion of their language input, which allows them to analyze better their language in its smallest component structures rather than memorizing larger, misleading chunks of input (e.g., Singleton & Newport, 2004). There is also evidence that the level of complexity of a visual stimulus can drive infant interest, with stimuli at either extreme (too simple or too complex) eliciting less interest from infants, a phenomenon that has been termed the *goldilocks effect* (Kidd, Piantadosi, & Aslin, 2012). These results and theories suggest that in some cases, too much complexity can be harmful to infant interest and learning.

The assumption that complexity is always harmful to learning may be erroneous, however. One issue is that complexity is in some ways a vacuous term; there are many different ways that something can be complex (e.g., number of items in a system, intricacies in spatial and temporal ordering between items), and it is frequently difficult to develop a principled way to define the term *a priori*. For example, any stimulus that proves to be difficult to learn can be described as complex in a post-hoc manner without any added explanatory value. To discuss complexity, then, we must establish a consistent and measurable operationalization. One possible instantiation of complexity is an increase in the amount of information (e.g., Shannon, 2001). Such an increase in information – such as in the case of the acoustic properties of real versus artificial speech – may not always be problematic for learning (e.g., Thiessen et al., 2005). For example, research indicates that adults may have a bias toward learning about inter-correlated features over single features. Specifically, Stadler (1992) found that sequences with higher levels of statistical structure were learned better than sequences with lower levels of statistical structure. Similarly, Billman (1989) reported that adults showed facilitation for rule learning in an implicit artificial language learning task when additional rules were correlated with the rules to be learned. Similar principles may be involved in infant learning. For example, young children can use distributional cues to learn grammatical gender categories but only when redundant cues signal category structure (Gerken, Wilson, & Lewis, 2005). Research on infant discovery of nonadjacent conditional relations between elements (e.g., *a-X-b*, where *a* predicts *b* and *X* is a variable element) demonstrates that greater variability in intervening elements (and thus greater complexity) actually helps infants to learn the nonadjacent regularity (Gómez, 2002). In particular, greater complexity may be helpful in situations where multiple sources of information converge (e.g., several imperfect acoustic and statistical cues all point to the same coherent units, or words, in speech).

Several studies of infant learning and complexity provide insight into this issue. One study tested whether infants exhibited better learning when the input contained multiple related regularities, or whether learning was superior when the input was simplified (Thiessen & Saffran, 2009). The authors found that increased complexity actually helped infants between 6.5 and 8 months of age, who learned melodies and lyrics more easily when they were presented in combination than when either was presentation in isolation. These results indicate that infants – like adults – sometimes benefit from greater complexity while learning about conditional relations between elements. These results parallel the findings of another study that found that for 6-month-old infants, visual speech cues enhanced phoneme discrimination, despite providing an additional source of information to be processed (Teinonen, Aslin, Alku, & Csibra, 2008). Also relevant is a study in which 12-month-olds were able to succeed on a statistical learning task in which the artificial language was characterized by multiple levels of regularities (Saffran & Wilson, 2003). Despite this added complexity, infants were able to segment multi-word utterances on the basis of statistical structure and subsequently discover word-order level regularities. Together, these studies question the assumption that increased complexity always impairs infant learning, and they suggest that the increased complexity of natural language may not pose an intractable challenge to infants' discovery of word forms via statistical cues. In particular, some aspects of the complexity that characterizes real language may actually facilitate word segmentation to the extent that multiple sources of information converge with each other and infants are able to take advantage of that convergence.

Similar principles regarding complexity can be seen in infant learning of form-based category abstraction (e.g., Gómez & Lakusta, 2004). In these paradigms, infants are familiarized with strings that conform to auditory structures that conform to the form *aX* and *bY* (e.g., *b* elements were paired with *Y* but not *X*). Typically, *a* and *b* elements consist of two distinct one syllable strings (e.g., the *a* element might consist of the words *alt* and *ush*). X and Y elements are drawn from larger pools (e.g., 6 distinct elements) and are distinguished by a phonological characteristic (number of syllables; e.g., *X*:

coomo; *kicey*; *Y*: *deech*; *tam*). Then, infants are tested on their ability to distinguish between novel strings that conform to the learned pattern and strings that violate the pattern (e.g., the *a* element is paired with either a one- or two-syllable novel word). By 12 months, when phonological cues (i.e., syllable length) are correlated with the distributional properties (i.e., the pattern of association between *a* and *X* and *b* and *Y* elements), infants can generalize these patterns to novel *X* and *Y* elements, discriminating legal from illegal strings after a brief (3 minute) training (e.g., Gómez & Lakusta, 2004; Lany & Gómez, 2008). In contrast, when phonological and distributional cues are uncorrelated (i.e., *X* and *Y* elements cannot be distinguished by syllable length), infants are unable to learn form-based categories. This paradigm was further extended Lany and Saffran (2010), who also first familiarized 22-month-old infants with strings in which phonological and statistical cues were either correlated or uncorrelated. Subsequently, infants in both conditions were trained on identical pairings between words and picture referents from two categories (animals or vehicles; e.g., *ong loga* was paired with an image of a meerkat). Only infants who had initially heard strings in which statistical and phonological cues were correlated were able to learn the associations between specific pairings, or show generalization to novel pairings. As this discussion indicates, complexity need not always hinder learning. The relation between complexity and learning outcomes appears to hinge on how the additional information in the input maps onto the underlying statistical structure. That said, it is important to acknowledge that much of the complexity inherent in language is likely to create challenges for infants who are acquiring language. More research is necessary, particularly with infants rather than adults (e.g., Frank et al., 2010, Vouloumanos, 2008; Vouloumanos & Werker, 2009) that explains infant acquisition of weaker or noisier statistical patterns.

Naturalism

A second feature of the early studies of statistical learning that contributed to their low ecological validity was the unnaturalness of the stimuli they used. As previously mentioned, this feature is intertwined with complexity insofar as language stimuli that are natural necessarily contain many sources of information in addition to the conditional relations between elements (e.g., real speech carries emotional information, words tend to co-occur with referents, and utterances contain both word- and phrase-level regularities). However, in this case, the concern is not about whether the learning mechanisms can scale up to greater complexity but rather that because the stimuli are highly artificial, infants do not approach the laboratory task in the same way that they approach the task of parsing real speech. Specifically, it may be that infants can succeed in these laboratory tasks, but they are not engaging in the same learning mechanisms that are at work in the discovery of real words because they do not process the artificial stimuli in the same way as real language. Although this sort of criticism is less common than concerns about the scalability of conclusions from artificial stimuli, both represent threats to ecological validity.

Several features of the traditional studies of statistical word segmentation contribute to their highly artificial nature. One such feature lies in the frequency and distribution of the words, which is problematic for the naturalism of the paradigm as well as the complexity, because natural input experienced by infants does not typically contain the amount of repetition as artificial language (although IDS does contain a level of repetition and redundancy that surpasses that of adult directed speech; Fernald, 2000). Perhaps the most critical difference between natural language and the artificial language stimuli used in statistical learning studies is the sheer amount of acoustic variation that is present in real language. This acoustic variation comes in many different forms. For example, coarticulation means that each phoneme in natural language differs slightly as a function of the phonemes that precede and follow it. In artificial language stimuli, this source of variability is not always included. A second source lies in the identity of the particular syllables that make up the utterances heard in both contexts (e.g., the syllable *da* appears in only one word in artificial language stimuli whereas in real language input *da* is a component syllable of multiple different words). Similarly, although word length varies in real language, the words used in stimuli have historically tended to be uniformly di- or tri-syllabic. Finally, whereas words in English are typically produced with lexical stress, most studies have used words that place equal stress on each syllable (but see Pelucchi et al., 2009; Thiessen & Erickson, 2013a; Thiessen & Saffran, 2003 for exceptions). Moreover, these words lacked indexical variation such as emotional valence, multi-talker variability, and the idiosyncratic variation that characterizes distinct tokens of produced over different time points. In addition, in contrast to the real language learning where words tend to co-occur temporally with referents, in these laboratory studies infants would typically hear speech without the opportunity to associate words with visual objects or events (but see Thiessen, 2010).

More recent research has made great strides in addressing these kinds of limitations, by designing experiments with increased ecological validity. These studies have done so by using stimuli for statistical segmentation paradigms that are simultaneously more complex and naturalistic (e.g., Hay, Pelucchi, Graf Estes, & Saffran, 2011; Pelucchi et al., 2009; Sahni et al., 2010). The artificial languages constructed for these studies better approximated the infant's real language input in several ways. Critical to the question of the specific role in statistical learning in word segmentation, these studies have involved modifications that increase the naturalism of the experimental input without introducing additional cues to word boundaries. One such modification is the use of speech produced by a human speaker rather than a speech synthesizer. As a result, the stimuli used in these studies contain many features of real speech. For example, in real speech words vary acoustically, depending on which words they precede and follow, because words cannot be produced by speakers without co-articulation. Similarly, these studies used stimuli that contained multiple tokens of particular words, meaning items exhibited slight acoustic variation. These and other modifications resulted in an increase in ecological validity.

Thiessen et al. (2005), one of the first studies to demonstrate infant segmentation with more complex and naturalistic stimuli, used a nonsense language that was more similar to the infant's real speech input in several ways. Whereas most studies of segmentation have used words of uniform length, typically two or three syllables (e.g., Thiessen & Saffran, 2003), Thiessen et al. (2005) used a language that varied the number of syllables in a word; the input contained both bi- and tri-syllabic nonsense words. In addition, although the only cue to word boundaries was the statistical structure of the speech, a speaker produced the words naturally in sentences. Because there were 12 distinct sentences, this meant that infants heard 12 distinct tokens of each word during the exposure phase. This contrasts with previous studies, in which each repetition of a particular word was created by splicing together acoustically identical instances of the word (e.g., Saffran, Aslin et al., 1996). Another by-product of using a language produced in sentences was that the input contained the characteristic intonation contours of real language. In one condition, the language was produced using the prosodic characteristics of *infant-directed speech* (IDS), which involves a characteristic exaggeration in pitch contour and elevation in pitch fundamental frequency. Performance in this condition was compared to a condition with a language produced with standard adult-directed speech (ADS). Infants who were exposed to the language produced in ADS did not show evidence of segmentation at test. Although the authors attributed this failure to segment to the brief exposure phase (in this study, infants heard each word only 12 or 24 times whereas the original studies used languages with 45 repetitions of each word), another possibility is that the failure to segment is the result of the use of a language that contained words of varying lengths, or a combination of the two factors. This possible explanation for their results accords with other studies that suggest infants have difficulty segmenting languages with varying word lengths, even with more word repetitions (Johnson & Tyler, 2010; Lew-Williams & Saffran, 2012), although with enough exposure it is possible that this difficulty may be overcome. However, in contrast to infants who listened to ADS, infants who heard the language produced in IDS discriminated words from lower-probability foils, indicating that they were able to segment the language. Thiessen et al. (2005) speculated that this IDS advantage for segmentation may lie in its capacity to hold attention (e.g., Werker, Pegg, & McLeod, 1994). Regardless of the specific mechanism, together these findings suggest that although increasing the variability of word lengths increases the difficulty of the segmentation task, this challenge can be mitigated by other features of the real input. In this case, IDS facilitated statistical word segmentation. IDS is pervasive cross-linguistically, and it has been argued to be a universal feature of communication with infants (e.g., Grieser & Kuhl, 1988). This modification of speech may play a critical role in the facilitation of word segmentation in the real world, where the input by nature contains noisy statistics.

Like Thiessen et al. (2005), other recent laboratory studies have used stimuli produced with IDS, as well as additional features contained in natural language that have furthered the trend of increasing ecological validity (e.g., Pelucchi et al., 2009). Pelucchi et al. (2009) tested the ability of the 8-month-old English-exposed infants to segment real language: the input comprised grammatically correct and semantically meaningful Italian sentences produced by a native speaker. This study contained many of the naturalistic features of the study by Thiessen et al. (2005), such as natural sentences produced in IDS with its characteristic intonation contours. These sentences included variable individual word tokens as well as words of variable length. Moreover, these real Italian sentences were produced using the typical rhythmic patterns of Italian. Italian shares the trochaic (strong–weak) lexical stress pattern of English, but the languages differ in many other dimensions such as phonotactics, allophonics, and other rhythmic aspects. This meant that the input was likely quite novel to the infants. Following a brief familiarization, infants were tested on their ability to discriminate high-probability items from low-probability items. At test, infants showed a significant preference to listen longer to high-probability items. Critically, both types of items were produced with trochaic stress. This meant that infants' preference could not be driven by an *a priori* preference for items with trochaic stress and were attributable to their successful segmentation of the speech stream.

Together, these studies take important first steps in increasing the ecological validity of laboratory studies of statistical learning. They demonstrate that in many cases the increased complexity that characterizes real language does not pose an insurmountable obstacle to infant statistical word segmentation. Moreover, in some cases, this complexity is actually beneficial to learning, such as in the case of a regular audiovisual association between words and referents (e.g., Thiessen, 2010). Critically, the types of heightened complexity that tend to either enhance learning are those that also increase the naturalism of the stimuli, which strengthens the argument that laboratory studies of statistical learning tap into the same processes that are at work. As noted previously (Saffran & Thiessen, 2003), such a pattern is unlikely to have occurred by chance. Instead, because languages are only able to survive when they contain the sorts of complexity that infants are able to learn, the link between complexity and naturalism likely evolved over our evolutionary history.

Summary

Studies of statistical word segmentation have made great strides in increasing their ecological validity since the pioneering studies of the 1990s. Despite the challenges inherent to the heightened complexity of the stimuli used in these studies, infants have been able to exhibit successful segmentation on the basis of statistical structure. However, practical issues in the design of the experimental paradigm preclude certain types of manipulations that might further bolster statistical accounts of word segmentation. For example, introducing the amount of noise that is present in the real world into conditional probabilities would necessitate the creation of experiments that persist for months or years. Practical considerations limit the feasibility of such studies, but support for statistical learning accounts of language acquisition can also be garnered from exploring additional predictions made by such theoretical accounts. One such account is the Extraction and Integration Framework (Thiessen et al., 2013), which is a potential mechanistic account of the processes by which statistical learning allows infants to discover word forms and adapt to the structure of their language.

In the next section, we will use this account to derive and explore testable predictions in the context of potential links between variation in performance tasks of statistical learning and variation in real language outcomes. For example, one key prediction made by statistical bootstrapping accounts of word segmentation is that reliance on statistical cues to word boundaries – which do not depend on language-specific properties – will precede reliance on language-specific acoustic cues (e.g., Thiessen & Erickson, 2013a). This is exactly what research has demonstrated: when statistical cues and stress cues are placed in conflict, English-exposed 7-month-olds segment on the basis of transitional probabilities and ignore stress cues. In contrast, 9-month-olds ignore transitional probabilities and segment units that conform to the predominant stress pattern of English (Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003; see Thiessen & Erickson, 2013b for a replication of segmentation based on statistical cues with 5-month-olds). These findings support the predictions made by statistical bootstrapping accounts, whereby early sensitivity to statistics is what allows infants to acquire language-specific knowledge of acoustic regularities (see also Sahni et al., 2010 for evidence that 9-month-olds can use syllable co-occurrence to discover a novel prosodic cue). However, accounts of word segmentation that emphasize an early reliance on acoustic cues to word boundaries provide a poorer fit to these findings. With regard to natural language acquisition, these results suggest that early learning should be slow, if infants are relying initially on statistical cues. Unlike laboratory studies, real language contains a rich array of cues to word boundaries. As infants identify more of these cues, word segmentation should become more rapid and efficient. This trajectory is consistent with empirical reports of early vocabulary development (Bates et al., 1988; Gopnik & Meltzoff, 1987; Lifter & Bloom, 1989).

Predictions of a statistical learning account of language acquisition

Since the advent of statistical learning research on infants, researchers have suggested a link between statistical learning and language development (e.g., Graf Estes, Evans, Alibali et al., 2007; Graf Estes, Evans, & Else-Quest, 2007; Romberg & Saffran, 2010; Saffran, Aslin et al., 1996; Thiessen & Saffran, 2003). One of the advantages of a theoretical framework like the Extraction and Integration account is that it yields testable predictions that make it possible to assess the validity of the supposed link between statistical learning and language outcomes. For example, this account holds that statistical learning is a mechanism that plays an important role in infants' discovery of words in fluent speech, as well as adapting to the predominant prosodic and phonological patterns of the native language. Such an account predicts that if statistical learning abilities are relevant to real language acquisition, individual differences in these abilities should be linked to real language outcomes, such that individuals who perform better on tasks of statistical learning should also achieve superior language learning and processing outcomes. Evaluating a link between individual differences in language outcome and learning ability is fundamental to any enterprise that claims a central role for statistical learning in language development. A second set of predictions relates to statistical learning abilities in clinical populations with impaired or atypical language development. Namely, if the Extraction and Integration account suggests that the processes of extraction and integration are critical for language development, we might predict that (1) these processes should be intact in typically developing populations with good language outcomes, and (2) one or both of these processes might be functioning atypically in populations with impaired or atypical language development. In the following sections, we will examine these predictions.

Statistical learning and building a lexicon

A key prediction of a statistical learning account of language acquisition is that the product of statistical learning – the representations that emerge as a function of learning – should be ready for incorporation into the developing linguistic system. With respect to word segmentation, this prediction can be formulated simply: the representations formed over statistical segmentation tasks should be word-like in nature. Critics of statistical learning approaches have argued that statistical learning involves calculations of transitional probabilities and does not involve the storage of frequent and coherent word-like units (e.g., Endress & Mehler, 2009). One approach to addressing this question has been to test the extent to which the output of statistical learning (i.e., words) shares features of real words (e.g., Erickson, Thiessen, & Graf Estes, 2014; Graf Estes, Evans, Alibali et al., 2007; Graf Estes, Evans, & Else-Quest, 2007; Mirman, Magnuson, Graf Estes, & Dixon, 2008; Saffran, 2001a). A variety of models of statistical learning suggest that this should be the case because they argue that statistical learning gives rise to chunks, or unitary representations of frequent and statistically coherent clusters of sounds, rather than merely producing computations which do not result in word-like items stored in memory (e.g., Perruchet & Vinter, 1998). These chunks compete with each other, so that frequent and statistically coherent items will tend to be more robustly represented than items that are less frequent and less statistically coherent. Consider the chunk *tyba*, formed from the end of the word *pretty* and the beginning of the word *baby*. Because this chunk is only likely to occur in the context of these two lexical items, it always receives competition from both items and its representation is pushed down relative to real world items, which necessarily compete with fewer items. Therefore, when infants segment speech on the basis of its statistical structure, statistical learning accounts predict that they should create word-like representations rather than merely high-probability sequences of sounds without relevance to their native language (e.g., Hay et al., 2011; Saffran, 2001a). That is to say, the output of statistical learning should be an integrated perceptual unit that is stored in memory. This unit would be highly familiar and available for subsequent processing or mapping to referents, and might contrast with a case where high probability transitions between sounds are familiar, but have not been stored in memory as a unitary chunk.

Research that used a diverse set of methodologies supports the prediction that word forms discovered via statistical cues exhibit word-like properties (Graf Estes, Evans, Alibali et al., 2007; Graf Estes, Evans, & Else-Quest, 2007; Hay et al., 2011; Saffran, 2001a). In one such study, Saffran (2001a) tested whether preferences for statistical nonsense words parsed from fluent speech (e.g., *tibudo*) differed as a function of the lexical context. These nonsense words were either embedded into a highly familiar English sentence frame (i.e., "I like my [*tibudo*]") or a nonsense sentence frame matched on several dimensions (i.e., "*Zy fike ny [tibudo*]"). Eight-month-old infants listened longer to items that were words in the language, but this preference held only when the words were presented in the context of a real English sentence. Listening times for infants who were exposed to these words in the context of nonsense word frames were equivalent, which suggested that infants were treating these items as potential English words. An account in which the output of statistical learning is merely a high-probability sequence of sounds does not make different predictions for particular sentential frames. Thus, the finding that infants only treated statistical words and foils differently when they occurred in meaningful linguistic contexts is unanticipated by such an account. One caveat is that an alternative explanation to these results is possible, namely, that differences in familiarity of the sentential frames may have driven the results. However, other studies that have asked whether the output of statistical learning is word-like using different methodologies provide indirect support for the idea that infants were treating these items as potential English words.

For example, Erickson et al. (2014) familiarized 8-month-old infants with an artificial language and subsequently gave them an object categorization task in which category exemplars were paired with linguistic associates. They found that the status of the linguistic associate with respect to the artificial language influenced whether infants were successful in categorizing. In particular, statistical words (items containing high probability syllable transitions) were found to facilitate object categorization relative to statistically incoherent foils (items containing one low probability syllable transitions). In the categorization literature, linguistic associates have been linked to superior categorization performance relative to nonlinguistic associates (e.g., Balaban & Waxman, 1997; Ferry, Hespos, & Waxman, 2010). Thus, statistically coherent items share properties of real words. Perhaps most compellingly, Graf Estes and colleagues (Graf Estes, Evans, Alibali et al., 2007; Graf Estes, Evans, & Else-Quest, 2007) used a word learning paradigm to investigate whether the process of statistically segmenting words from fluent speech is related to that of mapping meanings to labels. They found that 17-month-old infants were able to map labels to objects when those labels are composed of novel syllable sequences with high-internal probabilities. In contrast, infants did not learn the mapping when the labels comprised familiar sequences with low internal probabilities. This indicates that the process of segmenting words from fluent speech is intimately linked to word learning, which is also a critical component of early language acquisition. Taken together, these studies suggest that infants treat high probability items – or statistical words – and thus the output of statistical learning, as possible English words rather than merely sound sequences with high internal probabilities. This strengthens the argument that statistical learning experiments tap into the mechanisms that play a role in real language acquisition, rather than merely indexing some artificial laboratory phenomenon with little relevance to developing a lexicon.

Individual differences

A related prediction is that if statistical word segmentation is important for language acquisition, it should be related to the differences in real language outcomes. There is evidence to suggest that large individual differences in language learning abilities exist, particularly when learning takes place in adulthood (e.g., Dörnyei, 2005; Johnson & Newport, 1989). Recently, a group of researchers have suggested that individual variation in language learning abilities may be linked to stable genetic differences (e.g., Wong, Morgan-Short, Ettlinger, & Zheng, 2012). For example, functional human polymorphisms that result in differing levels of prefrontal dopamine activity are linked to individual differences in cognitive processes such as procedural learning, working memory, and attention (e.g., Papaleo, Erickson, Liu, Chen, & Weinberger, 2012; Wong et al., 2012). The same kinds of individual differences may also influence statistical learning, and, through statistical learning, language outcomes. For example, dopamine is known to influence the function of the basal ganglia, hippocampus, and prefrontal cortex, three regions that are likely to play a role in at least some aspects of statistical learning (e.g., Karuza et al., 2013; McNealy et al., 2006; for reviews for the dopaminergic system, see Seamans & Yang, 2004; Shohamy & Adcock, 2010). To the extent that these regions are implicated in statistical learning, and influenced by individual differences in dopaminergic functions, a logical conclusion is that individual differences in language acquisition linked to differences in dopamine may be driven, in part, by individual differences in statistical learning. Given the number of neural learning systems that may play a role in language learning (e.g., striatal regions; Karuza et al., 2013; Lim, Fiez, & Holt, 2014; Tricomi, Delgado, McCandliss, McClelland, & Fiez, 2006), this kind of individual variation may have a wide variety of underlying causes, and potentially different effects on statistical learning processes and outcomes. An investigation of stable genetic differences and language proficiency will likely prove valuable to our understanding of individual variability in language learning abilities.

Whereas much of the research on variability in language outcomes has focused on adults, this variability may be present in infancy. Indeed, a statistical learning account of individual differences predicts that at least a portion of this variability is rooted in individual differences in using statistical information to break into language. Infants who have trouble segmenting words from fluent speech may not have many word forms stored in memory to associate with real world referents, and thus may have smaller vocabularies than infants who are adept at segmenting speech. In addition, these infants may show additional disadvantages such as slower comprehension in online sentence processing. In contrast, infants who are skilled at extracting word forms from speech may have greater prospects to learn about the meaning and other aspects of these extracted word forms (e.g., Graf Estes, 2015). This argument is related to an argument made by Fernald and Marchman (2012), namely that individual differences in processing speed in infancy constitutes meaningful variability that shows continuity over time, and is related to real language outcomes such as vocabulary size and growth trajectories. Using longitudinal methods, they demonstrated that speed and accuracy in early understanding of language predicts cognitive and language outcomes at 8 years of age (Fernald, Perfors, & Marchman, 2006; Marchman & Fernald, 2008; see also Tsao, Liu, & Kuhl, 2004 for evidence that speech perception in infancy is related to language development in the second year of life). Longitudinal data linking statistical word segmentation abilities in infancy to later language proficiency outcomes are important to determine the role that statistical learning plays in the acquisition of natural languages and would strengthen statistical accounts of word discovery such as the Extraction and Integration Framework (e.g., Arciuli & Torkildsen, 2012; Thiessen et al., 2013).

Relatively few studies have investigated the possibility of a link between word segmentation and later language development. One exception is an experiment by Newman, Ratner, Jusczyk, Jusczyk, and Dow (2006) that tested infants between 7.5 and 12 months of age on their ability to recognize words from passages across a change in speaker's gender and their ability to use phonotactic cues to segment. They investigated whether these skills were related to their expressive vocabulary size at 2 years (Newman et al., 2006). They found that children who had larger vocabularies at 2 years had generally outperformed their peers on the segmentation tasks in the second half of their 1st year of life. In a subsequent analysis, they found that although segmenters and nonsegmenters differed in their later vocabulary size, they did not differ in general cognitive abilities as measured by the Kaufman Brief Intelligence test (K-BIT; Kaufman, 1990). This finding is problematic for the potential alternative explanation that children who were simply smarter or possessed faster processing speed were better equipped to learn words. Instead, this finding is consistent with the idea that it was their superior segmentation skills that led to their larger vocabularies, with two caveats. First, while it is important to establish that links between segmentation and later language acquisition are not entirely mediated by general cognitive abilities, this should not be taken as evidence that there is no relation between statistical learning and general cognitive abilities. Several theoretical accounts of statistical learning (including our own) posit just such a relationship. Second, these experiments looked at segmentation as a general construct, rather than specifically examining sensitivity to statistical structure. As such, it is premature to conclude the existence of a link between statistical learning abilities and language outcome, although these results are consistent with such a link. The relationship between early word segmentation and later vocabulary size established by Newman et al. (2006) has been since replicated by a recent prospective longitudinal study that examined the association between segmentation abilities at 7.5 months and vocabulary size at 2 years (Singh, Reznick, & Xuehua, 2012). In this study, infants were familiarized with words in isolation and subsequently tested on their ability to recognize those words in fluent speech. They found that segmentation performance was related to productive vocabulary size at 2 years of age. In addition, both vocabulary size and segmentation performance were related to general cognitive skills, as measured by the Bayley Scales of Infant Development (2nd ed.; BSID-II; Bayley, 1993).

Both Newman et al. (2006) and Singh et al. (2012) assessed a link between word segmentation and subsequent vocabulary size, without specifically assessing statistical learning abilities. We suggest that this correlation is due, at least in part, to the fact that infants who are better at segmenting are likely to be better statistical learners. However, an alternative possibility is that infants who are better at segmenting succeed for reasons that are unrelated to statistical learning. As such, it is important to directly investigate the link between statistical learning abilities and subsequent language outcomes. A growing number of studies have looked at the direct links between individual differences in a variety of statistical learning tasks and language proficiency. Evans, Saffran, and Robe-Torres (2009) tested the statistical word segmentation abilities of elementary school-aged children (between 6- and 14-years-olds). The authors found that statistical word segmentation was positively correlated with both receptive and expressive vocabulary. Similarly, another study found that adult implicit learning performance as measured by the Serial Reaction Time Task was correlated with the performance on two foreign language examinations (Kaufman et al., 2010). Another relevant study explored infants learning of form-based categories when statistical cues (i.e., distributional and phonological cues about label identity) were probabilistically predictive of semantic category membership (i.e., whether the label tended to be paired with animals or vehicles; Lany, 2014). Only the infants in the study with higher levels of grammatical development were able to use the statistical cues to support learning mappings between the labels and semantic categories, which is consistent with the idea that infants' sensitivity to relations between forms and meanings supports both learning of words and grammar. Finally, Misyak and Christiansen (2012) found that performance on tasks of auditory Artificial Grammar Learning – both of adjacent and nonadjacent regularities – was correlated with online sentence comprehension using a self-paced reading paradigm above, and beyond the contributions of factors such as working and short-term memory, vocabulary, reading experience, motivation, and fluid intelligence. Moreover, performance on the task of adjacent and nonadjacent statistical learning predicted processing of sentences characterized by short- and long-distance dependencies, respectively.

Together, these studies provide support for the idea that statistical learning and word segmentation performance are critically related to later vocabulary size, a real language outcome that is important for both communication and literacy (e.g., Anderson & Freebody, 1981). This correlation is consistent with the hypothesis that statistical learning plays a critical role in word segmentation. If statistical learning plays a fundamental role in word segmentation, then it follows that such a mechanism is important to developing a lexicon. Furthermore, the finding that statistical learning of nonadjacent relations predicted online processing of long-distance dependencies in sentences provides support for the idea that statistical learning is also important for syntax acquisition (e.g., Gomez & Gerken, 1999; Thompson & Newport, 2007). Critically, in several studies that collected measures of general intelligence and other cognitive factors, statistical learning and segmentation were found to explain variance in real language outcomes even after controlling for those factors (Misyak & Christiansen, 2012; Newman et al., 2006; see also Conway, Bauernschmidt, Huang, & Pisoni, 2010; Kaufman et al., 2010). If the predictive power of statistical learning abilities in accounting for variance in real language outcomes could be entirely accounted for by consideration of intelligence and other cognitive factors, this would be damaging to any account that posits a central role for statistical learning abilities in ultimate language attainment. That said, we do not mean to imply that statistical learning is likely to be completely independent of such factors. On the contrary, the relations between statistical learning and general cognitive abilities are likely to exist, although they may be complex and multi-faceted. For example, vocabulary knowledge constitutes an important part of intelligence testing. Similarly, if statistical segmentation involves the extraction of chunks from the input, it may be related to the storage aspects of working memory.[4] In turn, working memory is related to fluid intelligence, although the aspects of working memory that have to do with executive attention and manipulation of information rather than storage are likely to be more strongly related to fluid intelligence (Engle, Tuholski, Laughlin, & Conway, 1999). Although much work must still be done to specify how statistical learning fits into the range of human cognitive abilities, because it is a logical possibility that general cognitive abilities could mediate effects of statistical learning ability on language outcomes entirely, approaches that rule out intelligence as an alternative causal pathway are potentially informative. That said, existing research suggests that fluid intelligence and statistical learning as indexed by a variety of implicit learning paradigms are either unrelated or only weakly related (e.g., Kaufman et al., 2010; McGeorge, Crawford, & Kelly, 1997; Misyak & Christiansen, 2012; Reber, Walkenfeld, & Hernstadt, 1991; but see Brooks, Kempe, & Sionov, 2006).

Another approach to understanding the contributions of individual differences in statistical learning ability to language outcomes is to assess learning of non-linguistic stimuli. This is important because it is possible that success in linguistic statistical learning tasks is caused by strong language skills (e.g., more precise representations), rather than better statistical learning skills leading to better language learning outcomes. Correlations between visual statistical learning and individuals differences in language ability suggest that if there are stable individual differences in the ability to detect environmental statistical regularities, these differences are unlikely to be specific to prior experience with linguistic or auditory information (e.g., Arciuli & Simpson, 2012a, 2012b; Frost, Siegelman, Narkiss, & Afek, 2013). Arciuli & Simpson (2012a, 2012b) tested adults and elementary school-aged children using a visual analog of the typical auditory statistical learning paradigm. They reported that performance was significantly related to reading ability, independently of age and grade. Reading ability is underpinned by vocabulary knowledge among other factors (e.g., working memory; e.g., Cain, Oakhill, & Bryant, 2004) and has been suggested to reflect an ability to detect statistical regularities (e.g., Arciuli, Monaghan, & Seva, 2010; Plaut, McClelland, Seidenberg, & Patterson, 1996). Visual statistical learning has similarly been linked to adult second language acquisition (Frost et al., 2013). Frost et al. (2013) found that native English speakers who were better able to learn conditional relations between shapes were generally better at picking up Hebrew word morphology. A pretest indicated that the sensitivity to statistical structure of the shapes was not related to measures of generalized intelligence and working memory. In a similar vein, performance on a task of visual sequence learning was found to predict grammatical knowledge in hearing-impaired children (Conway, Pisoni, Anaya, Karpicke, & Henning, 2011). Kidd (2012) reported that implicit visual sequence learning but not explicit word pair learning predicted performance on a syntactic priming task. Finally, Shafto, Conway, Field, and Houston (2012) recently reported a positive correlation between visual statistical learning and receptive vocabulary in 8.5-month-olds. The correlation did not reach significance at a measurement five months later, although it was in the predicted direction. Thus, individual differences in statistical learning abilities within the visual modality are predictive of real language outcomes.

The studies described all relied on correlational methods, which cannot rule out the possibility that third variables are responsible for these effects, or indicate the directional of causality. Thus, caution must be observed in interpretation. However, the aggregate findings of these studies are consistent with the possibility that there exists a domain-general capacity to absorb statistical regularities in the input that uniquely accounts for variance in language outcomes. The existence of such a capacity fits with domain-general approaches that claim that language has phylogenetic and ontogenetic roots in basic perceptual and cognitive mechanisms rather than specialized learning mechanisms (e.g., Bates & MacWhinney, 1982; Tomasello, 2003), and it is consistent with an account in which infants higher in this ability show advantages at both the segmentation of speech and the acquisition of syntax. These putative advantages may result in a variety of superior language outcomes, such as larger vocabularies and superior acquisition and processing of syntactic structure (e.g., Kidd, 2012; Evans et al., 2009).

[4] Controversy exists regarding the distinction between working and short-term memory (e.g., Engle et al., 1999). Here, we use the term working memory as an umbrella term and distinguish between different aspects of working memory by referring to storage vs. executive attention.

However, much remains unknown about such a domain-general statistical learning capacity, particularly with respect to how it is related to other domain general cognitive abilities, many of which are at least moderately related to each other (e.g., executive functions and intelligence; Duncan, Emslie, Williams, Johnson, & Freer, 1996; Friedman et al., 2006). Indeed, there are theoretical reasons to believe that a general statistical learning ability should be related to other domain general cognitive skills. For example, the Extraction and Integration Framework suggests that the process of extraction should be related to attention and working memory, and the process of integration should be related to encoding specificity and prototype abstraction. Also uncertain is the extent to which different measures of statistical and implicit learning (e.g., the Serial Reaction Time task; Nissen & Bullemer, 1987; the Artificial Grammar Learning paradigm; Reber, 1967, 1969, 1989) are interrelated. Few studies have investigated the question of how multiple measures of statistical learning are related to each other and to other cognitive abilities (but see Gebauer & Mackintosh, 2007; Kaufman et al., 2010; Misyak & Christiansen, 2012). For example, although research indicates that attention is necessary for statistical learning (e.g., Baker et al., 2004; Toro et al., 2005), to our knowledge no study has investigated whether individual differences in statistical learning are related to individual differences in attention.

Relation to clinical phenotypes

Beyond studies that reveal the presence of significant correlations between statistical learning performance and real language outcomes in typically developing populations, studies of clinical populations may provide further insight into the prediction that statistical learning is related to building a lexicon and other language processes. Although the etiology of clinical disorders is typically complex and multifaceted, a potential avenue of inquiry is to determine whether clinical populations that display linguistic deficits or delays show disrupted statistical learning abilities. The mere presence of deficits in such populations cannot conclusively demonstrate that these linguistic difficulties stem from problems with statistical learning and not more global cognitive processing deficits; however, the existence of such deficits would align with an account in which some aspects of language difficulties may be related to disrupted statistical learning abilities. Further, if these populations show completely intact and typical statistical learning abilities in the face of language difficulties, it might suggest that other cognitive and perceptual abilities are more important to typical language functioning than statistical learning. Critically, we do not mean to imply that all cases of atypical language development can be explained by deficits in statistical learning. For example, the failure to acquire spoken language by children who are profoundly deaf is certainly not attributable to impaired statistical learning. However, there may be some cases of atypical language outcomes in which statistical learning does not function normally. In this section, we will examine the statistical learning abilities of selected clinical phenotypes characterized by atypical language development.

Specific manifestations of language impairments in different clinical populations might be produced by distinct patterns of disruptions to the machinery responsible for sensitivity to statistical structure. For example, the process of extracting frequent and statistically coherent clusters might be disrupted, such that word forms are never extracted, and thus never stored in memory. Another possible pattern of deficits might involve intact extraction and storing of word forms, but difficulty with other aspects of the learning process. For example, individuals might exhibit difficulties associating those stored word forms to individual items or categories of items in the world. Alternatively (or additionally), learners might experience problems with the process of integrating across the word forms to discover language-specific phonological patterns, which in turn might reduce efficiency of the later extraction of appropriate units from speech. Learners might also experience difficulties discovering the units over which to integrate. Alternatively, an infant might show relatively normal performance on these aspects of word learning, yet show difficulty integrating multiple sources of information when they point to different speech parsings. Finally, a disruption in statistical learning may also lead to deficits in producing and comprehending syntax among individuals from clinical populations. To the extent that the ability to acquire nonadjacent statistical regularities as measured by Artificial Grammar Learning experiments (e.g., Gómez, 2002; Reber, 1967, 1969, 1989) is related to the acquisition and processing of syntax, populations that exhibit difficulties in processing syntax may also be impaired at acquiring nonadjacent statistical regularities in laboratory studies.

Here, we explore the prediction that statistical learning may function atypically in a clinical population characterized by deficits in language: children with specific language impairment. As noted previously, the etiology and manifestation of these disorders is undoubtedly multifaceted, the result of pleiotropy and gene by gene interactions. Thus, we do not mean to imply that a deficit in statistical learning abilities is the primary causal factor in the etiology of developmental language disorders. Instead, we mean to use an examination of these clinical populations as a mean of examining the link between statistical learning and language. If, as we claim, statistical learning plays an important role in language development, at least some of the populations that have difficulties with language should show deficits in statistical learning.

Specific language impairment

Specific language impairment (SLI) is a disorder that affects between 3% and 10% of children (Tomblin et al., 1997). Children with SLI exhibit pervasive speech and language deficits such as delayed vocabulary and syntax acquisition (e.g., Lum & Bleses, 2012; Rice, Tomblin, Hoffman, Richman, & Marquis, 2004; van Der Lely & Battell, 2003). Although the name SLI implies linguistic specificity and considerable research has focused on deficits in phonological processing and verbal short-term memory abilities (e.g., Gathercole & Baddeley, 1990; Graf Estes, Evans, & Else-Quest, 2007), growing evidence casts doubt on the linguistic specificity of the disorder (e.g., Evans et al., 2009). Unlike disorders such as Williams syndrome or Down syndrome, which possess a clear etiology, SLI exhibits great heterogeneity (e.g., Leonard, 1998). This heterogeneity in manifestation and severity of impairments results from the fact that SLI is diagnosed when a child's oral language lags behind other areas of development for reasons that cannot be ascertained, such as when a child exhibits normal hearing, normal nonverbal intelligence test scores, and a lack of neurological damage (Leonard, 1998). As a result, caution must be observed with drawing conclusions about the nature and etiology of SLI; however, one alternative to the linguistic specificity hypothesis is that there is a general perceptual or cognitive locus to the disorder (e.g., Ullman & Pierpont, 2005).

This hypothesis is supported by studies that indicate that many individual with SLI possess at least one nonlinguistic deficit, such as impairments in motor functions and rapid temporal processing (e.g., Alcock, Passingham, Watkins, & Vargha-Khadem, 2000; Botting & Conti-Ramsden, 2001; Dewey & Wall, 1997; Evans et al., 2009; Fazio, 1996; Gathercole & Baddeley, 1990; Johnston & Ramstad, 1983; Montgomery, 1993; Noterdaeme, Mildenberger, Minow, & Amorosa, 2002; Tallal & Piercy, 1973; Tallal et al., 1996; Vargha-Khadem, Watkins, Alcock, Fletcher, & Passingham, 1995; for a review, see Bishop, 1992). In particular, many of the impairments involve complex sequences of movement, such as oromotor movements, or moving pegs (e.g., Alcock et al., 2000; Bishop, 2002). Schwartz and Regan (1996) found a strong correlation between auditory language comprehension and performance on fine motor tasks, which may suggest that these procedural learning deficits are tightly linked to the language difficulties. Implicit learning and statistical learning are believed to be closely related learning mechanisms, if they are not the same mechanism (e.g., Perruchet & Pacton, 2006).[5] If procedural learning deficits are in fact a core component of SLI, it will remain to be seen whether such deficits are a causal factor in the manifestation of the disorder, or merely a superficial by-product of atypicalities in phonological or perceptual processing. Regardless of the etiology, statistical learning accounts predict that individuals with SLI, who have atypical language outcomes, will also exhibit atypicalities in extraction, in integration, or in both statistical learning processes. In addition, the Extraction and Integration Framework differs from other accounts insofar as it predicts that if atypicalities in extraction exist, they should be causally related to deficits in working memory. In contrast, other accounts have argued that if working deficits in SLI exist, they should be considered peripheral rather than core deficits (Lum et al., 2012).

Several studies indicate that extraction, or sensitivity to conditional statistical structure, is impaired in SLI (e.g., Lum, Gelgic, & Conti-Ramsden, 2010; Tomblin, Mainela-Arnold, & Zhang, 2007), which

[5] Other perspectives regard different implicit learning tasks as engaging distinct learning mechanisms (e.g., Gebauer & Mackintosh, 2007) although these issues are complicated by the possibility that task reliability may play a role in obscuring inter-relations between implicit learning tasks (e.g., Salthouse, McGuthry, & Hambrick, 1999).

may be related to both the small vocabularies found in SLI and difficulties individuals with SLI experience acquiring syntactic patterns. Evans et al. (2009) tested the ability of elementary school-aged children with SLI to segment speech on the basis of statistical structure. Whereas typically developing children matched on age and nonverbal IQ exhibited above chance performance and a positive correlation with vocabulary size after 21 minutes of exposure to the language, children with SLI required a much longer exposure to the language to show above-chance performance and the positive correlation between ability and vocabulary size demonstrated by the typically developing children using a shorter exposure phase. Moreover, when the input consisted of tones characterized by the same statistical structure as the speech, children with SLI failed to show evidence of learning even after 42 minutes of exposure, despite the fact that typically developing children showed above chance performance. These results indicate that statistical learning abilities are atypical in this clinical population, and this impairment is not limited to linguistic materials. Children with SLI may be impaired at the process of statistical learning whereby frequent and statistically coherent units are culled from speech.

If the ability to benefit from the presence of statistical structure is causally linked to vocabulary acquisition, this disruption in the ability to extract frequent and coherent units from speech may play a role in the language deficits displayed by children with SLI. Specifically, if the segmentation of speech on the basis of statistical structure requires more time for infants who will later be diagnosed with SLI, they might demonstrate a considerably delayed trajectory in the discovery of word forms from speech. In turn, having relatively few word forms stored in memory means that there will be fewer word forms to associate with real world referents. Additionally, the Extraction and Integration account suggests that to the extent word–object association is facilitated by statistical learning (extracting and binding word–object pairings), then children who have difficulty with the process of extraction may also be impaired in associating lexical forms with referents. Evidence is mixed with regard to this latter prediction, perhaps due to heterogeneity in the population of children diagnosed with SLI (e.g., Lum et al., 2010). Regardless, our account predicts that difficulty with extraction may prevent children with SLI from developing typically-sized vocabularies, which can subsequently lead to further language-related deficits (e.g., Bailey & Snowling, 2002). In contrast to traditional accounts of SLI invoking language-specific impairments, our perspective suggests that difficulties with lexical acquisition in SLI arise, at least in part, from a more general deficit in statistical learning abilities (e.g., Evans et al., 2009; Hsu, Tomblin, & Christiansen, 2014).

Another hallmark of SLI, in addition to lexical deficits (e.g., Gray, 2005), is impaired acquisition of syntax (e.g., Bailey & Snowling, 2002; Gopnik & Crago, 1991; Hansson & Nettelbladt, 1995; Rothweiler & Clahsen, 1993). Syntax involves the acquisition of nonadjacent regularities in addition to adjacent regularities and has thus been posited to be related to the statistical learning of nonadjacent regularities (e.g., Gomez & Gerken, 1999). This proposal is consistent with several studies that have found links between sensitivity to statistical structure in Artificial Grammar Learning and procedural learning tasks and acquisition and processing of syntactic structure in typically developing individuals (e.g., Kidd, 2012; Misyak & Christiansen, 2012; Lum et al., 2014). According to the Extraction and Integration Framework, both extraction and integration may underlie the learning of nonadjacent regularities and consequently syntax. Thus, pervasive difficulties in learning the statistical structure of input through extraction as well as integration may be related to the syntactic deficits found in SLI. Additionally, the Extraction and Integration Framework predicts that differential impairment to extraction and integration should lead to somewhat different outcomes: whereas vocabulary acquisition is more dependent on the process of extraction, we suggest that the acquisition of syntax is relatively more dependent on the process of integration.

Although a comprehensive assessment of different patterns of statistical learning difficulties in children with SLI is not yet available, some evidence exists that is consistent with this prediction. Tomblin et al. (2007) tested the implicit learning abilities of both typically developing adolescents and adolescents with SLI using the Serial Reaction Time Task, which consists of long sequences and thus may involve the detection of both adjacent and nonadjacent regularities. Although both groups improved over time, indicating sensitivity to the sequences, the rate of learning was slower for adolescents with SLI. Moreover, this difference in learning rate was found when language impairment was defined in terms of grammatical impairments, but not when language impairment was defined in terms of vocabulary group differences. This may account for some of the individual variability in the manifestation of SLI, as many individuals with SLI are likely to show slightly different constellations of impairments as a result of the way individuals who experience language difficulties are grouped. Lum et al. (2010) also found that children with SLI were impaired relative to typically developing peers on the Serial Reaction Time Task. In addition, they found that children with SLI showed impairments on a task of explicit learning. Specifically, children with SLI experienced difficulty with a paired associate task, but only when the elements to be associated were verbal rather than visual. The authors interpreted these findings as evidence that although some children with SLI have difficulties with declarative memory, others have intact declarative memory, at least in the context of auditory information (e.g., Leonard et al., 1982; Whitehurst, Novak, & Zorn, 1972). In contrast, the implicit learning of sequences may be impaired in a greater percentage of individuals with SLI. This dissociation between lexical/declarative and procedural abilities may provide a partial explanation of why some children with SLI eventually appear to catch up with their peers, if these children are able to compensate for poor implicit learning with an overreliance on lexical memory and the application of explicit rules to memorize complex grammatical forms (e.g., Gopnik & Crago, 1991; Ullman & Gopnik, 1999).

Much remains unknown regarding the nature of SLI as a result of both the diagnostic criteria and the uncertainty as to which phenotypic characteristics reflect causes and which reflect relatively more superficial by-products of other impairments. In spite of the uncertainty surrounding the etiology of SLI, the few studies that exist are consistent with the prediction that individuals with SLI will show impairments on tasks of statistical learning (e.g., Evans et al., 2009; Hsu, Tomblin, & Christiansen, 2014; Tomblin et al., 2007). Moreover, the Extraction and Integration Framework makes novel predictions with regard to the etiology of SLI that may prompt future investigation. Prior accounts of the link between implicit learning and SLI have suggest that the heterogeneity in outcomes is due to the fact that some children with SLI show “true” procedural learning deficits (i.e, procedural but not declarative impairment), while others show “peripheral” deficits unrelated to statistical learning (Lum et al., 2012). In contrast, our framework conceptualizes this heterogeneity in terms of differential impairments to the two underlying processes necessary for statistical learning, extraction and integration. This leads to testable predictions regarding the patterns of impairments in individuals with SLI. For example, according to the Extraction and Integration Framework, a subset of individuals with SLI may show deficits in the detection of conditional regularities (e.g., statistical segmentation) that are coupled to atypicalities in working memory and attention. In contrast, other perspectives would predict that even if working memory and attentional deficits are found in individuals with SLI, they should not constitute a core aspect of the impairment (Lum et al., 2012). Another subset of individuals with SLI may show impairments in a variety of tasks that rely on integration, or inducing central tendency from a group of exemplars (e.g., Artificial Grammar Learning). Different combinations of impairments in these processes may contribute to the heterogeneity in the manifestation of SLI. More studies are needed to fully characterize the nature of impairments in statistical learning, and to test whether only sensitivity to conditional regularities is impaired or whether integrating over exemplars to form prototypes also shows deficits. In addition, such investigations should consider the role of modality. Whereas the Extraction and Integration Framework proposes that statistical learning tasks may be best grouped according to their relative reliance on the processes of extraction and integration, other perspectives emphasize the importance of modality in grouping tasks (e.g., Conway & Christiansen, 2006). Consequently, future research should test which of these organizations has more explanatory power in the context of SLI.

Summary

Together, these studies are consistent with the possibility that deficits in conditional and distributional statistical learning may exist in clinical populations characterized by language deficits or delays. An important caveat is that the etiology of clinical disorders is complex and multi-faceted. Thus, although deficits in statistical learning may be a causal factor in the language impairments displayed by certain clinical populations, this is not the only viable possibility. For example, statistical learning deficits may be one of the multiple causal factors, or the result of a lower level perceptual or phonological deficit. That said, the results from the reviewed studies are largely consistent with a story in which difficulties in adaptation to the statistical structure of the environment play an important role in the segmentation of speech to acquire a lexicon, the acquisition of syntactic structure, and integrating across stored exemplars to induce knowledge of language-specific phonological patterns. Rather than being restricted to linguistic materials, these impairments can be seen in nonlinguistic domains as well and likely contribute to the poor language outcomes associated with these clinical phenotypes. Regardless of whether statistical learning turns out to be a central causal factor or a relatively more peripheral indicator of other causes, the reviewed studies support the prediction made by statistical accounts of language acquisition such the Extraction and Integration Framework that populations with atypical language outcomes will show atypical statistical learning.

Conclusion

The evidence outlined so far – research on statistical learning using more ecologically valid stimuli as well as correlations between individual variability and real language outcomes – supports the claim that statistical learning is related to the acquisition of language in naturalistic settings. With respect to statistical word segmentation, support for the predictions made by statistical bootstrapping accounts of word segmentation lies in studies that employ increased ecological validity of the stimuli via heightened complexity and naturalism (e.g., Hay et al., 2011; Pelucchi et al., 2009; Thiessen et al., 2005). These studies suggest that, in many cases, the complexity of natural language may not pose an intractable obstacle for statistical learning. They suggest that this mechanism is sufficiently powerful to play a fundamental role in developing a lexicon. This lexicon then allows infants to adapt further to the phonological patterns of their native language (i.e., patterns that require an understanding of how that phonological pattern is correlated with word position, as in the case of lexical stress or phonotactics) via distributional statistical learning, which then increases the speed and accuracy of subsequent utterance segmentations. For example, knowing that stressed syllables tend to initiate words in English allows a listener to correctly segment an utterance after a single hearing (e.g., Cutler & Carter, 1987). English-speaking infants and adults are more accurate at segmenting a speech stream with iambic stress following training with lists of isolated iambic words (e.g., Thiessen & Erickson, 2013a; Thiessen & Saffran, 2003). Familiarity with words themselves also aids segmentation, as illustrated by the finding that infants can segment words that follow familiar words (e.g., *mommy*; Bortfeld et al., 2005). Similarly, statistical learning may provide an opportunity to acquire phrase level regularities by extracting and storing multi-word chunks (e.g., Saffran, 2001b; Saffran & Wilson, 2003). Integrating over the chunks provides an opportunity to identify consistent syntactic frames (e.g., *I'm VERBing it*).

Further support for relevance of statistical learning to real language acquisition comes from studies that correlate performance on tasks of conditional statistical learning with performance on meaningful measures of real language outcomes such as vocabulary size, reading comprehension, and syntax acquisition (Arciuli & Simpson, 2012a; Evans et al., 2009; Kidd et al., 2012). Although correlational methods cannot rule out the possibility that general cognitive abilities are at the core of all of these tasks, in cases where measures of general cognitive abilities or intelligence are reported, statistical learning and segmentation abilities explain a unique portion of the variance. Thus statistical learning is likely to be important for language acquisition even beyond word segmentation, the focus of the earliest studies of this flexible domain-general mechanism (e.g., Saffran, Aslin et al., 1996). More research is necessary to characterize the nature of potentially causal links between statistical learning and later language outcomes. For example, longitudinal research demonstrating links between early statistical learning abilities and later vocabulary outcomes would be of value. Moreover, research on clinical populations with language difficulties or delays shows that statistical learning and word segmentation abilities are impaired relative to those possessed by typically developing populations (e.g., Evans et al., 2009). If statistical learning does indeed play a critical role in the discovery of words and the acquisition of syntactic and phonological regularities, disruptions to particular aspects of statistical learning mechanisms may be a causal factor in some of the language delays and deficits present in these populations. Data from clinical populations have the potential to help us develop a better understanding of the process of statistical learning, by virtue of presenting an opportunity to examine how patterns of impairments in specific disorders map on to disruptions in different aspects of the statistical learning process.

Comparison to other accounts

The Extraction and Integration Framework is not the first attempt to explain how statistical learning contributes to language acquisition. It does, however, differ from prior accounts on several dimensions. First and foremost, it attempts to unify discussion of statistical learning of sequential elements (e.g., syllables in an utterance) with discussion of learning of distributional patterns (e.g., prototypically arrangement of phonemes in phonotactic learning). This is in contrast to many models of word segmentation that suggest that tracking of sequential probabilities arises from a completely different set of processes than learning of phonological regularities such as predominant lexical stress, phonology, or phonotactics. For example, in the StaGE model (Adriaans & Kager, 2010), learners identify words via calculation of transitional probabilities and then deduce phonological and phonotactic constraints via a hierarchical ranking of constraints similar to that of Optimality Theory (Tesar & Smolensky, 2000). Similarly, Mersad and Nazzi (2011) proposed that learners use a hierarchical ranking of phonological cues to parse the speech stream, relying on sequential statistics only in cases where these cues are uninformative. In addition to invoking very different kinds of learning mechanisms and representations than the Extraction and Integration Framework, both of these accounts explain segmentation in terms of a hierarchy of cues, such that higher-ranking cues should always outrank lower-ranking cues. By contrast, our account proposes that phonological cues exert an influence on segmentation as a function of guiding attention on the basis of the similarity between the current input and (representations of) prior input. From this perspective, the "ranking" of cues may change in a fluid way as a function of changes in the distributions to which children are exposed, or the similarity on a variety of features between the current input and prior input.

Another difference between our account, and prior theoretical accounts of statistical learning, is in the nature of the learning mechanisms we invoke. One aspect of this difference is in terms of the kinds of computational principles that are thought to give rise to sensitivity to statistical structure in the input. Many prior models of statistical learning have suggested some form of explicit computation of, or sensitivity to, transitional probabilities (e.g., Adriaans & Kager, 2010; Frank et al., 2010). By contrast, from our perspective, learners may appear to be sensitive to transitional probabilities because they are storing chunks of the input, which – due to basic memory processes such as interference – are biased toward statistically coherent chunks over the course of exposure to the input (e.g., Perruchet & Vinter, 1998). This difference leads these different classes of models to be sensitive to different kinds of manipulations of the input. For example, consider a hypothetical artificial language in which *A* is followed by *B* 50% of the time, and by *C* 50% of the time. By contrast, consider a hypothetical grammar in which *A* is followed by *B* 50% of the time, and also be elements *C*, *D*, *E*, *F*, and *G*, each of which occurs 10% of the time. From the perspective of computing transitional probabilities, the relation between *A* and *B* is identical across both languages, and so learning of the *A*–*B* relationship should proceed equivalently. From an extraction perspective, however, the *A*–*B* relationship has a much stronger competitor in the first language than in the second language, so the identical 50% transitional probability should be learned differently across the two languages.

A final notable dimension of difference between our account and others is in the mechanistic, biological processes that learners use (as opposed to the formal description of the computations) to achieve sensitivity to statistical structure. We have discussed several such processes over the course of describing our framework, including activation, interference, and decay. However, other process claims are possible. For example, our framework suggests that statistical learning shares deep commonalities with implicit learning, as it invokes chunking as an explanation for sensitivity to conditional statistical structure (Perruchet & Vinter, 1998). Chunking has frequently been invoked as an explanatory construct in a wide variety of implicit learning tasks, from Artificial Grammar Learning to Serial Reaction Time tasks (e.g., Koch & Hoffmann, 2000; Reber, 1969). The claim that statistical learning and implicit learning share underlying processes (and may even be different labels for the same process) is not a universally acknowledged one; for example, it is in opposition to accounts of statistical learning that rely on computation of transitional probabilities (e.g., Adriaans & Kager, 2010). As such, the relationship between statistical and implicit learning is still debatable and an important topic for future research (e.g., Conway & Christiansen, 2006; Hamrick & Rebuschat, 2012; Perruchet & Pacton, 2006).

Even among "implicit-style" accounts of statistical learning, the Extraction and Integration Framework differs from other accounts in its claims about the processes underlying learning. One such difference is that our framework suggests a unified set of processes operating across domains. Many other implicit learning accounts – and statistical learning accounts – have suggested that different processes underlie learning in different modalities (e.g., Conway & Christiansen, 2006). An alternative "division of labor" approach to implicit learning is Ullman's suggestion that arbitrary associations (of the form necessary to acquire word forms and meaning) are accomplished by different processes than the learning of procedural tasks (e.g., Ullman, 2004, 2005; Ullman & Pierpont, 2005). Similarly, many accounts of language acquisition that depend on implicit learning draw a sharp distinction between procedural and declarative knowledge (e.g., Ullman, 2001). By contrast, our approach suggests that explicit knowledge emerges from implicit learning processes: for example, the knowledge that the word *dog* means "4-legged animal that barks," arises from repeated exposure to pairings of lexical form and object, and integration across those exposures that allows for a more abstract interpretation to emerge.

Consistent with these differences in claims about the process of learning, our perspective differs from other accounts in terms of the neurological underpinning of statistical learning. We have highlighted two systems: a prefrontally-mediated attentional system necessary for extraction and a hippocampally- and cortically-mediated system necessary for integration. Other accounts, particularly implicit learning accounts and traditional multiple memory systems accounts (e.g., Squire, 1987, 1992; Ullman, 2004), have argued that the hippocampus is unnecessary for implicit learning and have instead focused on the role of the basal ganglia, and particularly the striatum, in tasks related to both extraction and integration (e.g., Lieberman, Chang, Chiao, Bookheimer, & Knowlton, 2004; Ullman, 2006). Indeed, many studies have demonstrated basal ganglia recruitment in implicit learning tasks (e.g., Lieberman et al., 2004; Ullman, 2006). However, some have questioned the traditional division of labor between explicit and implicit memory systems on the grounds that clean distinctions often break down (e.g., Shohamy & Turk-Browne, 2013). Moreover, evidence from patient data is mixed regarding the necessity of these particular brain regions for various implicit learning tasks (e.g., Knowlton et al., 1992; Knowlton, Mangels, & Squire, 1996; Schapiro et al., 2014; Smith, Siegert, McDowall, & Abernethy, 2001), and fMRI studies frequently report both basal ganglia and medial temporal lobe activation for both declarative and procedural tasks. (e.g., Degonda et al., 2005; see also Frank, Loughry, & O'Reilly, 2001; McNab & Klingberg, 2008 for suggestions that the basal ganglia are engaged in working memory, a process we hypothesize to be related to extraction). Thus, much uncertainty remains regarding the neural systems underlying the various tasks that have been termed statistical learning. It is also entirely possible that these tasks do not recruit the same mechanisms (e.g., notably, as mentioned previously the Serial Reaction Time Task differs from statistical word segmentation and Artificial Grammar Learning tasks insofar as it uses a motor response which may provide participants with a source of implicit feedback, which is known to influence learning; e.g., Lim et al., 2014). Further work will be needed to disentangle the computational and neural mechanisms that support performance in these tasks.

Future directions

The literature reviewed here highlights the nuanced and multi-faceted nature of the broad variety of phenomena that have been termed "statistical learning" following early studies of word segmentation. According to the Extraction and Integration Framework, statistical learning refers to adaptation to statistical regularities in the environment that relies on two main processes and their interactions: extraction and integration. Underlying these processes are general properties of the human cognitive architecture such as attention, as well as working memory and long term memory. This framework makes novel testable predictions about how individual variability in statistical learning ability might cluster according to whether the tasks tap into extraction or integration, or a combination of the two processes. As we have discussed above, this account yields predictions at several levels of analysis – behavioral, computational, and biological – that differ from other accounts of the role of statistical learning in language acquisition, and from more traditional accounts of linguistic learning emphasizing the role of symbolic computations (e.g., Pinker, 1998; Pinker & Ullman, 2002). As such, an important avenue for future research is to test these predictions and use them to refine (or disconfirm) the framework.

In conclusion, statistical learning is a powerful mechanism that infants can use to adapt to various aspects of the statistical structure of their native language. Going beyond traditional conceptualizations of statistical learning as a transitional probability calculator, we argue that statistical learning is fundamentally two complementary processes: the extraction of coherent units and integration across those units to induce further structure (e.g., segmenting trochaic units from speech and integrating across those stored units to learn the correlation between lexical stress and word position). After structure has been induced, it can bias attention to support further learning (e.g., segmenting additional trochaic units from speech). This conceptualization of statistical learning is able to account for a wide range of phenomena than typically explained (e.g., Goldwater, Griffiths, & Johnson, 2009; Perruchet & Vinter, 1998; but for notable exceptions, see Adriaans & Kager, 2010; Thiessen & Erickson, 2013b; Thiessen et al., 2013). This perspective leads to a set of predictions in computational, behavioral, and clinical domains that stem from the conceptualization of sensitivity to statistical regularities in different domains as the same underlying processes rooted in features of human perception and memory.

Although the evidence considered here supports the importance of statistical learning in language acquisition phenomena, many questions remain unanswered. The field would benefit from comprehensive longitudinal investigations of how capacities for statistical learning in infancy are related to language proficiency over the course of subsequent years. Similarly, there is a need for research programs that test the limits of statistical learning in situations where the statistics more closely approximate the noise found in real language. Relatedly, research on the integration of statistical cues with other cues to word boundaries (i.e., acoustic regularities) must move beyond investigations of the relative strength of particular cues into research programs that describe how multiple sources of information are integrated to yield segmented word forms. In addition to behavioral research, neuroimaging, computational modeling and corpus analyses have the potential to inform our understanding of language acquisition, through the precise specification of the underlying processes. Despite the clear gaps in our knowledge, and the many questions yet to be addressed, the use of statistical learning to provide a unified account of many disparate phenomena shows promise in enhancing our understanding of language development.

Acknowledgments

This research was supported by a grant from the National Science Foundation awarded to E.D.T. (BCS0642415). This material is based on work supported by the National Science Foundation under Grant No. (0946825) awarded to L.C.E. We would like to thank Anna Fisher, Lori Holt, David Rakison, Michael Kaschak, Matt Lehet, Kevin Jarbo, Casey Roark, and anonymous reviewers for their comments on drafts of this manuscript.

References

Abbot-Smith, K., Lieven, E., & Tomasello, M. (2001). What preschool children do and do not do with ungrammatical word orders. *Cognitive Development*, *16*(2), 679–692.

Adriaans, F., & Kager, R. (2010). Adding generalization to statistical learning: The induction of phonotactics from continuous speech. *Journal of Memory and Language*, *62*(3), 311–331.

Alcock, K. J., Passingham, R. E., Watkins, K. E., & Vargha-Khadem, F. (2000). Oral dyspraxia in inherited speech and language impairment and acquired dysphasia. *Brain and Language*, *75*(1), 17–33.

Aly, M., & Turk-Browne, N. B. (2015). Attention stabilizes representations in the human hippocampus. *Cerebral Cortex* (in press).

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). Newark, DE: International Reading Association.

Anglin, J. M., Miller, G. A., & Wakefield, P. C. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development*.

Arciuli, J., Monaghan, P., & Seva, N. (2010). Learning to assign lexical stress during reading aloud: Corpus, behavioral, and computational investigations. *Journal of Memory and Language*, *63*(2), 180–196.

Arciuli, J., & Simpson, I. C. (2012a). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, *36*, 286–304.

Arciuli, J., & Simpson, I. C. (2012b). Statistical learning is lasting and consistent over time. *Neuroscience Letters*, *517*(2), 133–135.

Arciuli, J., & Torkildsen, J. V. K. (2012). Advancing our understanding of the link between statistical learning and language acquisition: The need for longitudinal data. *Frontiers in Language Sciences*, *3*, 324.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324.
Bailey, P. J., & Snowling, M. J. (2002). Auditory processing and the development of language and literacy. *British Medical Bulletin*, *63*(1), 135–146.
Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science*, *15*(7), 460–466.
Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology*, *64*(1), 3–26.
Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*(3), 1382–1407.
Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, *19*(3), 241–248.
Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar*. Cambridge, UK: Cambridge University Press.
Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In *Language acquisition: The state of the art* (pp. 173–218).
Bayley, N. (1993). *Bayley scales of infant development* (2nd ed.). San Antonio, TX: The Psychological Corporation.
Bertoncini, J., & Mehler, J. (1981). Syllables as units in infant speech perception. *Infant Behavior and Development*, *4*, 247–260.
Bijeljac-Babic, R., Bertoncini, J., & Mehler, J. (1993). How do 4-day-old infants categorize multisyllabic utterances? *Developmental Psychology*, *29*(4), 711.
Billman, D. (1989). Systems of correlations in rule and category learning: Use of structured input in learning syntactic categories. *Language and Cognitive Processes*, *4*(2), 127–155.
Bishop, D. V. (1992). The underlying nature of specific language impairment. *Journal of Child Psychology and Psychiatry*, *33*(1), 3–66.
Bishop, D. V. (2002). Motor immaturity and specific speech and language impairment: Evidence for a common genetic basis. *American Journal of Medical Genetics*, *114*(1), 56–63.
Bomba, P. C., & Siqueland, E. R. (1983). The nature and structure of infant form categories. *Journal of Experimental Child Psychology*, *35*(2), 294–328.
Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*(4), 298–304.
Botting, N., & Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment (SLI). *International Journal of Language & Communication Disorders*, *36*(4), 421–432.
Boucher, L., & Dienes, Z. (2003). Two ways of learning associations. *Cognitive Science*, *27*(6), 807–842.
Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*(1), 93–125.
Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33–B44.
Brooks, P. J., Kempe, V., & Sionov, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied Psycholinguistics*, *27*(02), 185–209.
Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127–132.
Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*(1), 31–42.
Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, *33*(2), 111–153.
Cartwright, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition*, *63*(2), 121–170.
Chomsky, N. (1959). A review of BF skinner's verbal behavior. *Language*, *35*(1), 26–58.
Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, *13*(2&3), 221–268.
Cole, R., & Jakimik, J. (1980). A model of speech perception. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 133–163). Hillsdale, NJ: Erlbaum.
Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*(3), 356–371.
Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(1), 24.
Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities pitting abstract against stimulus-specific representations. *Psychological Science*, *17*(10), 905–912.
Conway, C. M., Pisoni, D. B., Anaya, E. M., Karpicke, J., & Henning, S. C. (2011). Implicit sequence learning in deaf children with cochlear implants. *Developmental Science*, *14*(1), 69–82.
Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*, 1119–1130.
Cutler, A. (1996). Prosody and the word boundary problem. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 87–100). Hillsdale, NJ: Erlbaum.
Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, *2*(3), 133–142.
Degonda, N., Mondadori, C. R., Bosshardt, S., Schmidt, C. F., Boesiger, P., Nitsch, R. M., et al. (2005). Implicit associative learning engages the hippocampus and interacts with explicit associative learning. *Neuron*, *46*(3), 505–520.
Dewey, D., & Wall, K. (1997). Praxis and memory deficits in language-impaired children. *Developmental Neuropsychology*, *13*(4), 507–512.
Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Erlbaum.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology*, *30*(3), 257–303.
Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.
Elman, J. L. (Ed.), (1998). *Rethinking innateness: A connectionist perspective on development* (Vol. 10). The MIT Press.
Emberson, L. L., Liu, R., & Zevin, J. D. (2013). Is statistical learning constrained by lower level perceptual organization? *Cognition*, *128*(1), 82–102.
Endress, A. D., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language*, *60*(3), 351–367.
Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology. General*, *128*(3), 309–331.
Erickson, L. C., Thiessen, E. D., & Graf Estes, K. (2014). Statistically coherent labels facilitate categorization in 8-month-olds. *Journal of Memory and Language*, *72*, 49–58.
Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *52*(2), 321–335.
Fazio, B. B. (1996). Serial memory in children with specific language impairment: Examining specific content areas for assessment and intervention. *Topics in Language Disorders*, *17*(1), 58–71.
Fernald, A. (2000). Speech to infants as hyperspeech: Knowledge-driven processes in early word recognition. *Phonetica*, *57*(2–4), 242–254.
Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development*, *83*(1), 203–222, 11.
Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*(1), 98–116.
Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3-and 4-month-old infants: An advantage of words over tones. *Child Development*, *81*(2), 472–479.
Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, *108*(2), 477–499.
Finn, A. S., Lee, T., Kraus, A., & Kam, C. L. H. (2014). When it hurts (and helps) to try: The role of effort in language learning. *PLoS ONE*, *9*(7), e101806.
Fiser, J. Z., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*, 499–504.
Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, *117*(2), 107–125.
Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(2), 137–160.
French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*(4), 614.
French, R. M., & Cottrell, G. W. (2014). *TRACX 2.0: A memory-based, biologically-plausible model of sequence segmentation and chunk extraction*. Poster presented at the 36th Annual Conference of the Cognitive Science Society. Quebec City, Canada.
Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*(2), 172–179.
Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125.
Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science*, *24*(7), 1243–1252.
Gambell, T., & Yang, C. (2004, August). *Statistics learning and universal grammar: Modeling word segmentation*. In First Workshop on Psycho–computational Models of Human Language Acquisition (p. 49).
Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, *29*(3), 336–360.
Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(1), 34–54.
Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, *32*(02), 249–268.
Gervain, J., & Guevara, R. (2012). The statistical signature of morphosyntax: A study of Hungarian and Italian infant-directed speech. *Cognition*, *125*(2), 263–287.
Giroux, I., & Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science*, *33*(2), 260–272.
Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436.
Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science*, *7*(5), 567–580.
Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, *7*(5), 219–224.
Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, *31*(3), 305–320.
Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition*, *112*(1), 21–54.
Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–135.
Gopnik, A., & Meltzoff, A. N. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. *Child Development*, *58*, 1523–1531.
Gopnik, M., & Crago, M. B. (1991). Familial aggregation of a developmental language disorder. *Cognition*, *39*(1), 1–50.
Graf Estes, K. (2015). How infants find words. In J. Taylor (Ed.), *The Oxford handbook of the word*. Oxford: Oxford University Press.

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*(3), 254–260.
Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *50*(1), 177–195.
Gray, S. (2005). Word learning by preschoolers with specific language impairment: Effect of phonological or semantic cues. *Journal of Speech, Language, and Hearing Research*, *48*(6), 1452–1467.
Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology*, *24*(1), 14.
Hamrick, P., & Rebuschat, P. (2012). How implicit is statistical learning. In *Statistical learning and language acquisition* (pp. 365–382).
Hansson, K., & Nettelbladt, U. (1995). Grammatical characteristics of Swedish children with SLI. *Journal of Speech, Language, and Hearing Research*, *38*(3), 589–598.
Harris, Z. S. (1954). Distributional structure. *Word*.
Harris, Z. S. (1955). From phoneme to morpheme. *Language*, *31*(2), 190–222.
Hay, J., 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.
Hayes, J. R., & Clark, H. H. (1970). Experiments in the segmentation of an artificial speech analog. In *Cognition and the development of language* (pp. 221–234). New York: Wiley.
Hayes, N. A., & Broadbent, D. E. (1988). Two modes of learning for interactive tasks. *Cognition*, *28*(3), 249–276.
Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*(2), 96–101.
Hsu, H. J., Tomblin, J. B., & Christiansen, M. H. (2014). Impaired statistical learning of non-adjacent dependencies in adolescents with specific language impairment. *Frontiers in Psychology*, 5.
Hunter, M. A., Ames, E. W., & Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. *Developmental Psychology*, *19*(3), 338–352.
Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*(4), 548–567.
Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science*, *12*(1), 131–141.
Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339–345.
Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99.
Johnston, J. R., & Ramstad, V. (1983). Cognitive development in pre-adolescent language impaired children. *International Journal of Language & Communication Disorders*, *18*(1), 49–55.
Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23.
Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development*, *64*, 675–687.
Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, *61*(8), 1465–1476.
Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*(3–4), 159–207.
Jusczyk, P. W., & Thompson, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Perception & Psychophysics*, *23*(2), 105–109.
Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language*, *127*(1), 46–54.
Kaschak, M. P., & Saffran, J. R. (2006). Idiomatic syntactic constructions and language learning. *Cognitive Science*, *30*(1), 43–63.
Kaufman, A. S. (1990). *Kaufman Brief Intelligence Test: KBIT*. AGS, American Guidance Service. Circle Pines, MN: AGS Publishing.
Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*(3), 321–340.
Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, *7*(5), e36399.
Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*(1), 171–184.
Kim, G., Lewis-Peacock, J. A., Norman, K. A., & Turk-Browne, N. B. (2014). Pruning of memories due to context-based prediction error. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 8997–9002.
Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters*, *461*(2), 145–149.
Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42.
Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*(5280), 1399–1402.
Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, *3*(3), 172–179.
Koch, I., & Hoffmann, J. (2000). Patterns, chunks, and hierarchies in serial reaction-time tasks. *Psychological Research*, *63*(1), 22–35.
Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews. Neuroscience*, *5*(11), 831–843.
Lany, J. (2014). Judging words by their covers and the company they keep: Probabilistic cues support word learning. *Child Development*, *85*(4), 1727–1739.
Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, *19*(12), 1247–1252.
Lany, J., & Saffran, J. R. (2010). From statistics to meaning infants' acquisition of lexical categories. *Psychological Science*, *21*(2), 284–291.

Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.
Leonard, L. B., Schwartz, R. G., Chapman, K., Rowan, L. E., Prelock, P. A., Terrell, B., et al. (1982). Early lexical acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *25*(4), 554–564.
Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science*, *14*(6), 1323–1329.
Lew-Williams, C., & Saffran, J. R. (2012). All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition*, *122*(2), 241–246.
Lieberman, M. D., Chang, G. Y., Chiao, J., Bookheimer, S. Y., & Knowlton, B. J. (2004). An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *Journal of Cognitive Neuroscience*, *16*(3), 427–438.
Lifter, K., & Bloom, L. (1989). Object knowledge and the emergence of language. *Infant Behavior and Development*, *12*, 395–423.
Lim, S. J., Fiez, J. A., & Holt, L. L. (2014). How may the basal ganglia contribute to auditory categorization and speech perception? *Frontiers in Neuroscience*, *8*, 230.
Lim, S. J., Holt, L. L., & Fiez, J. A. (2013). *Context-dependent modulation of striatal systems during incidental auditory category learning*. Poster Presented at the Annual Meeting of the Society for Neuroscience, San Diego, CA.
Lum, J. A., & Bleses, D. (2012). Declarative and procedural memory in Danish speaking children with specific language impairment. *Journal of Communication Disorders*, *45*(1), 46–58.
Lum, J. A., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. T. (2014). Procedural learning deficits in specific language impairment (SLI): A meta-analysis of serial reaction time task performance. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *51*, 1–10.
Lum, J. A., Conti-Ramsden, G., Page, D., & Ullman, M. T. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *48*(9), 1138–1154.
Lum, J. A., Gelgic, C., & Conti-Ramsden, G. (2010). Procedural and declarative memory in children with and without specific language impairment. *International Journal of Language & Communication Disorders*, *45*(1), 96–107.
Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes*, *66*(3), 309–332.
Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(3), F9–F16.
Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London B: Biological Sciences*, *204*(1156), 301–328.
Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465–494.
Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology. General*, *134*(4), 477–500.
Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111.
McCall, R. B., & Kagan, J. (1970). Individual differences in the infant's distribution of attention to stimulus discrepancy. *Developmental Psychology*, *2*(1), 90–98.
McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457.
McGeorge, P., Crawford, J. R., & Kelly, S. W. (1997). The relationships between psychometric intelligence and learning in an explicit and an implicit task. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *23*(1), 239–245.
McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*(1), 103–107.
McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: Neural mechanisms underlying speech parsing. *The Journal of Neuroscience*, *26*(29), 7629–7639.
Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143–178.
Mersad, K., & Nazzi, T. (2011). Transitional probabilities and positional frequency phonotactics in a hierarchical model of speech segmentation. *Memory & Cognition*, *39*(6), 1085–1093.
Miller, G. A., & Gildea, P. M. (1987). How children learn words. *Scientific American*, *257*(3), 94–99.
Mirman, D., Graf Estes, K., & Magnuson, J. S. (2010). Computational modeling of statistical learning: Effects of transitional probability versus frequency and links to word learning. *Infancy*, *15*(5), 471–486.
Mirman, D., Magnuson, J. S., Graf Estes, K., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, *108*(1), 271–280.
Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*(1), 302–331.
Montgomery, J. W. (1993). Haptic recognition of children with specific language impairment: Effects of response modality. *Journal of Speech, Language, and Hearing Research*, *36*(1), 98–104.
Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language*, *35*(5), 666–688.
Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, *66*(4), 911–936.
Musz, E., Weber, M. J., & Thompson-Schill, S. L. (2015). Visual statistical learning is not reliably modulated by selective attention to isolated events. *Attention, Perception, & Psychophysics*, *77*(1), 78–96.
Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology. Human Perception and Performance*, *24*(3), 756–766.
Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology*, *42*(4), 643–655.
Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*(1), 11–28.

Newport, E. L., & Aslin, R. N. (2000). *Innately constrained learning: Blending old and new approaches to language acquisition*. In Proceedings of the 24th Annual Boston University Conference on Language Development (Vol. 1).
Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of nonadjacent dependencies. *Cognitive Psychology*, *48*, 127–162.
Ngon, C., Martin, A., Dupoux, E., Cabrol, D., Dutat, M., & Peperkamp, S. (2013). (Non) words,(non) words,(non) words: Evidence for a protolexicon during the first year of life. *Developmental Science*, *16*(1), 24–34.
Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32.
Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395.
Noterdaeme, M., Mildenberger, K., Minow, F., & Amorosa, H. (2002). Evaluation of neuromotor deficits in children with autism and children with a specific speech and language disorder. *European Child & Adolescent Psychiatry*, *11*(5), 219–225.
Onnis, L., & Thiessen, E. (2013). Language experience changes subsequent learning. *Cognition*, *126*, 168–284.
Onnis, L., Waterfall, H. R., & Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition*, *109*(3), 423–430.
Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(7), 2745–2750.
Papaleo, F., Erickson, L., Liu, G., Chen, J., & Weinberger, D. R. (2012). Effects of sex and COMT genotype on environmentally modulated cognitive control in mice. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(49), 20160–20165.
Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–685.
Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, *21*(1), 178–185.
Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238.
Perruchet, P., & Tillmann, B. (2010). Exploiting multiple sources of information in learning an artificial language: Human data and modeling. *Cognitive Science*, *34*(2), 255–285.
Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246–263.
Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2–3), 115–154.
Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics*, *34*(4), 516–530.
Pinker, S. (1998). Words and rules. *Lingua*, *106*(1), 219–242.
Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*(11), 456–463.
Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115.
Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3), 353–363.
Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science*, *275*(5306), 1604–1610.
Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863.
Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, *81*(1), 115–119.
Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology. General*, *118*(3), 219–235.
Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *17*, 888–896.
Reich, P. A. (1986). *Language development*. Englewood Cliffs, NJ: Prentice-Hall.
Rice, M. L., Tomblin, J. B., Hoffman, L., Richman, W. A., & Marquis, J. (2004). Grammatical tense deficits in children with SLI and nonspecific language impairment: Relationships with nonverbal IQ over time. *Journal of Speech, Language, and Hearing Research*, *47*(4), 816–834.
Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews*: *Cognitive Science*, *1*(6), 906–914.
Rothweiler, M., & Clahsen, H. (1993). Dissociations in SLI children's inflectional systems: A study of participle inflection and subject-verb-agreement. *Logopedics, Phoniatrics, Vocology*, *18*(4), 169–179.
Saffran, J. R. (2001a). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, *81*(2), 149–169.
Saffran, J. R. (2001b). The use of predictive dependencies in language learning. *Journal of Memory and Language*, *44*(4), 493–515.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.
Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.
Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*, 101–105.
Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, *39*(3), 484–494.
Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy*, *4*(2), 273–284.
Sahni, S. D., Seidenberg, M. S., & Saffran, J. R. (2010). Connecting cues: Overlapping regularities support cue discovery in infancy. *Child Development*, *81*(3), 727–736.
Salthouse, T. A., McGuthry, K. E., & Hambrick, D. Z. (1999). A framework for analyzing and interpreting differential aging patterns: Application to three measures of implicit learning. *Aging, Neuropsychology, and Cognition*, *6*(1), 1–18.
Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *Journal of Cognitive Neuroscience*, *26*(8), 1736–1747.
Schwartz, M., & Regan, V. (1996). Sequencing, timing, and rate relationships between language and motor skill in children with receptive language delay. *Developmental Neuropsychology*, *12*(3), 255–270.

Seamans, J. K., & Yang, C. R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Progress in Neurobiology*, *74*(1), 1–58.
Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, *275*(5306), 1599–1603.
Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *16*(4), 592–608.
Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, *5*(1), 3–55.
Shafto, C. L., Conway, C. M., Field, S. L., & Houston, D. M. (2012). Visual sequence learning in infancy: Domain-general and domain-specific associations with language. *Infancy*, *17*(3), 247–271.
Shohamy, D., & Adcock, R. A. (2010). Dopamine and adaptive memory. *Trends in Cognitive Sciences*, *14*(10), 464–472.
Shohamy, D., & Turk-Browne, N. B. (2013). Mechanisms for widespread hippocampal involvement in cognition. *Journal of Experimental Psychology. General*, *142*(4), 1159–1170.
Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*(1), 1–32.
Singh, L., Reznick, J. S., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: A longitudinal analysis. *Developmental Science*, *15*(4), 482–495.
Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology*, *49*(4), 370–407.
Slone, L. K., & Johnson, S. P. (2015). *Statistical and chunking processes in infants' and adults' visual statistical learning*. Poster presented at the biennial meeting of the Society for Research in Child Development, Philadelphia, PA.
Smith, J., Siegert, R. J., McDowall, J., & Abernethy, D. (2001). Preserved implicit learning on both the serial reaction time task and artificial grammar in patients with Parkinson's disease. *Brain and Cognition*, *45*(3), 378–391.
Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, 549–565.
Soto, D., Mäntylä, T., & Silvanto, J. (2011). Working memory without consciousness. *Current Biology*, *21*(22), R912–R913.
Squire, L. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience*, *4*(3), 232–243.
Squire, L. R. (1987). *Memory and the brain*. New York: Oxford University Press.
Stadler, M. A. (1992). Statistical structure and implicit serial learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *18*(2), 318–327.
Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*(1), 80–132.
Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., et al. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science*, *271*(5245), 81–84.
Tallal, P., & Piercy, M. (1973). Defects of non-verbal auditory perception in children with developmental aphasia. *Nature*, *241*(5390), 468–469.
Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, *108*(3), 850–855.
Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*(1), 21.
Tesar, B., & Smolensky, P. (2000). *Learnability in optimality theory*. Cambridge, MA: MIT Press.
Thiessen, E. D. (2010). Effects of visual information on adults' and infants' auditory statistical learning. *Cognitive Science*, *34*(6), 1093–1106.
Thiessen, E. D. (2011). Domain general constraints on statistical learning. *Child Development*, *82*(2), 462–470.
Thiessen, E. D., & Erickson, L. C. (2013a). Discovering words in fluent speech: The contribution of two kinds of statistical information. *Frontiers in Psychology*, *3*, 590.
Thiessen, E. D., & Erickson, L. C. (2013b). Beyond word segmentation: A two-process account of statistical learning. *Current Directions in Psychological Science*, *22*(3), 239–243.
Thiessen, E. D., & Erickson, L. C. (2015). The statistical approach to word segmentation. In T. H. Mintz (Ed.), *Current trends in statistical approaches to language acquisition*. New York: Psychology Press.
Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant directed speech facilitates word segmentation. *Infancy*, *7*, 49–67.
Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*, *139*(4), 792–814.
Thiessen, E. D., & Pavlik, P. I. (2013). iMinerva: A mathematical model of distributional statistical learning. *Cognitive Science*, *37*(2), 310–343.
Thiessen, E. D., & Pavlik, P. I. (2015). *Modeling the effects of phonemic distribution on children's use of phonemic contrasts*. (in press).
Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7-to 9-month–old infants. *Developmental Psychology*, *39*(4), 706–716.
Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development*, *3*(1), 73–100.
Thiessen, E. D., & Saffran, J. R. (2009). How the melody facilitates the message and vice versa in infant learning and memory. *Annals of the New York Academy of Sciences*, *1169*(1), 225–233.
Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, *3*(1), 1–42.
Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, *11*(1/2), 61–82.
Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge MA: Harvard University Press.
Tomblin, J. B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural learning in adolescents with and without specific language impairment. *Language Learning and Development*, *3*(4), 269–293.
Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, *40*(6), 1245–1260.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25–B34.
Tricomi, E., Delgado, M. R., McCandliss, B. D., McClelland, J. L., & Fiez, J. A. (2006). Performance feedback drives caudate activation in a phonological learning task. *Journal of Cognitive Neuroscience*, *18*(6), 1029–1043.
Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*, *75*(4), 1067–1084.
Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(02), 105–122.
Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1), 231–270.
Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory, and practice* (pp. 141–178). Washington, DC: Georgetown University Press.
Ullman, M. T. (2006). Is Broca's area part of a basal ganglia thalamocortical circuit? *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *42*(4), 480–485.
Ullman, M. T., & Gopnik, M. (1999). Inflectional morphology in a family with inherited specific language impairment. *Applied Psycholinguistics*, *20*(1), 51–117.
Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *41*(3), 399–433.
Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038–1066.
van Der Lely, H. K., & Battell, J. (2003). Wh-movement in children with grammatical SLI: A test of the RDDR hypothesis. *Language*, 153–181.
Van de Weijer, J. (2001). *The importance of single-word utterances for early word recognition*. In Proceedings of ELA 2001 Lyon, France.
Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P., & Passingham, R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(3), 930–933.
Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*(2), 729–742.
Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology*, *45*(6), 1611–1617.
Werker, J. F., Pegg, J. E., & McLeod, P. J. (1994). A cross-language investigation of infant preference for infant-directed communication. *Infant Behavior and Development*, *17*(3), 323–333.
Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, *103*(1), 147–162.
Werker, J., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.
Whitehurst, G. J., Novak, G., & Zorn, G. A. (1972). Delayed speech studied in the home. *Developmental Psychology*, *7*(2), 169–177.
Wong, P., Morgan-Short, K., Ettlinger, M., & Zheng, J. (2012). Linking neurogenetics and individual differences in language learning: The dopamine hypothesis. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *48*(9), 1091–1102.
Yang, C. D. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences*, *8*(10), 451–456.
Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. *Language Learning and Development*, *4*(1), 32–62.
Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420.
Yurovsky, D., Fricker, D. C., Yu, C., & Smith, L. B. (2013). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin & Review*, 1–22.
Yurovsky, D., Yu, C., & Smith, L. B. (2012). Statistical speech segmentation and word learning in parallel: Scaffolding from child-directed speech. *Frontiers in Psychology*, *3*, 374.