4

STATISTICAL LEARNING APPROACHES TO STUDYING LANGUAGE DEVELOPMENT

Ryan A. Cannistraci, Rodrigo Dal Ben, Ferhat Karaman, Sara Parvanezadeh Esfahani, and Jessica F. Hay

Introduction

Anyone who has spent time around young children will notice that they acquire language at an alarming rate – they go from being non-verbal in early infancy to speaking in full and grammatically complex sentences between their second and third birthdays. A question that has been of particular interest to researchers is, how do children learn language so quickly? One prominent hypothesis is that infants may begin to learn language by picking up on the patterns or structures in their linguistic environment even before they know very much about the specific characteristics of their native language – a capability that is often referred to as statistical learning.

The idea that infants can track regularities in their linguistic input was popularized in 1996 by Jenny Saffran, Richard Aslin, and Elissa Newport. They showed that following 2 minutes of familiarization with an artificial language (e.g., *pabikugolapitimore* . . .), 8-month-old infants could discriminate syllable sequences that had occurred together consistently (i.e., words) from those that spanned word boundaries (i.e., part-words). These findings suggested that infants may be able to track the co-occurrence patterns (also referred to as sequential statistics, conditional statistics, and transitional probability) between syllables to discover words in continuous speech (Saffran, Aslin, & Newport, 1996). This work has been replicated many times with other artificial (e.g., Aslin, Saffran, & Newport, 1998; Thiessen & Saffran, 2003) and natural languages (e.g., Karaman & Hay, 2018; Pelucchi, Hay, & Saffran, 2009a, b) (see Section 2 of this volume for more information about statistical learning and speech segmentation; see the Conclusions section of this chapter for extensions of this work beyond speech and humans), and such tracking is believed to occur implicitly or incidentally (i.e., without conscious awareness; e.g., Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). Finding words in continuous speech is just one of the many learning challenges infants face when acquiring language. Infants also need to discover which sounds are relevant in their native language, how sound sequences map to meaning, and how elements in language are organized relative to one another, to name a few. The past two decades of research have suggested that infants have remarkable computational abilities that may allow them to track the types of statistical regularities relevant to each of these learning challenges.

Through the use of highly controlled experiments, laboratory demonstrations of statistical learning have been very informative regarding infants' capabilities. Yet we still do not know if, and to what degree, infants rely on these computational abilities to learn language outside of the laboratory. We suggest that the field may be able to glean insight into the role of statistical learning by looking at how individual differences in statistical learning abilities relate to language development.

This chapter is organized into four sections. In each section we will begin by highlighting a language learning challenge faced by infants: 1. discovering the sounds in language, 2. finding words in continuous speech, 3. mapping words to meaning, and 4. learning rudimentary grammar. In each section we will provide an overview of the kind of statistical regularities available in the input and review a representative sampling of the evidence that suggests that infants can track those statistics (for a recent review, see also Saffran & Kirkham, 2018). For each of these learning challenges, if relevant data are available, we will discuss what is known about how individual differences in statistical learning are related to language development in both typical and atypical infant and child populations. Our goal in this chapter is to give the reader a flavor of the ways that statistical learning in infancy (and childhood) is studied in the lab and to highlight evidence to suggest that it may account for some meaningful individual differences in language acquisition.

Discovering the sounds in language

One of the major learning challenges faced by infants early in development is discovering which speech sounds and speech sound variations are relevant in their native language. Infants demonstrate early sensitivity to many of the sounds found across the world's languages (Kuhl, Tsao, Liu, Zhang, & Boer, 2001). As they gain experience with their native language, infants tend to show reduced sensitivity to many non-native phonemes (Lalonde & Werker, 1995; Werker & Tees, 1984) and show an increased ability to be able to discriminate native-language phonemes (Kuhl, Stevens, Hayashi, Deguchi, Kiritani, & Iverson, 2006; Narayan, Werker, & Beddor, 2010). This process has been referred to as perceptual narrowing (for a review, see Maurer & Werker, 2014) or native-language attunement (see Chapter 15 for a more in-depth discussion of the development of speech perception). Evidence of perceptual narrowing in monolinguals is typically observed around 10 to 12 months of age for consonants (e.g., Kuhl et al., 2001; Werker & Tees, 1984), and somewhat earlier for vowels (e.g., Polka & Werker, 1994). For infants being exposed to more than one language, there is the added challenge of learning two phonological systems, and bilinguals tend to show somewhat protracted perceptual narrowing (see Chapter 16 for further information about bilingual language acquisition).

The speed with which infants learn to attend to some sound contrasts and ignore others has been of interest to researchers for a long time. One of the most prominent hypotheses is that infants home in on the sound patterns of their native language(s) by tracking statistical regularities in the frequencies and distributions of sounds in acoustic space (Maye, Werker, & Gerken, 2002; Werker, Yeung, & Yoshida, 2012). In the extant literature, this type of pattern detection is referred to as distributional learning, and is arguably one of lowest level forms of statistical learning (see Ambridge, Kidd, Rowland, & Theakston, 2015). Consistent with this idea, each of the world's languages offers the developing infant different distributions of acoustic information. So, for example, American English /r/ and /l/, which are lexically contrastive in that they can be used to differentiate word meaning (e.g., *rate* vs. *late*), are produced with a bimodal distribution along a particular acoustic dimension (i.e., the frequency of the third formant (F3)). That is, in natural speech, exemplars of /r/ cluster together at lower values of F3, whereas exemplars of /l/ cluster together at higher values of F3, with only a small degree of acoustic overlap between productions of /r/ and /l/. In contrast, in natural Japanese productions there is a unimodal cluster of sounds, called a flap, that overlaps with American English /r/ and /l/ (see Figure 4.1 for an

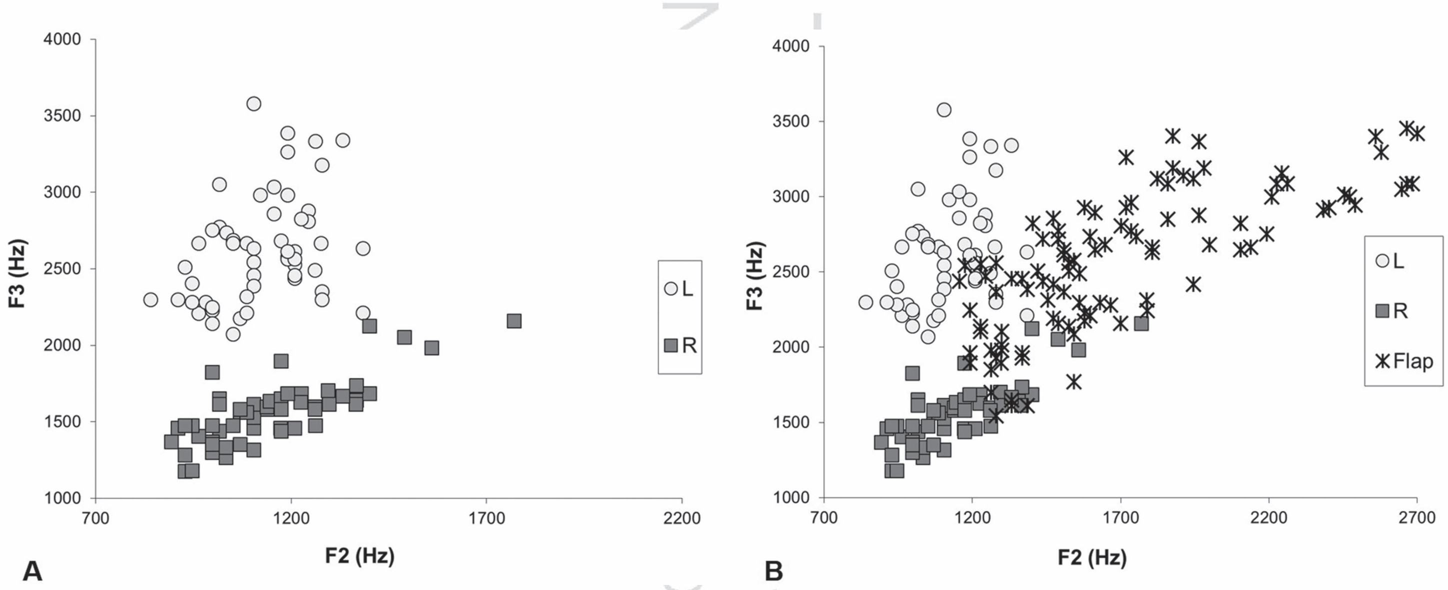

Figure 4.1 Comparison of English speakers' productions for /l/ and /r/ (a) overlaid with Japanese speakers' production of the flap consonant (b).

Source: Lotto, Sato, & Diehl (2004).

example of natural productions of English /l/ and /r/, and the Japanese flap, from Lotto, Sato, & Diehl, 2004). Although both English- and Japanese-learning infants are initially able to discriminate English /r/ from /l/, by the time they are 11 months old Japanese-learning infants have collapsed the English contrast into a single category, consistent with the distributional information found in their language (Kuhl et al., 2001). Thus, sensitivity to this type of statistical distributional information might account for perceptual narrowing observed in the first year of life.

Maye, Werker, and Gerken (2002) tested the prediction that infants are sensitive to distributional information by presenting 6- and 8-month-old infants with a bimodal or unimodal distribution during familiarization and then testing their subsequent speech perception abilities. They created a continuum of eight tokens between English /da/ and /ta/ (voiceless unaspirated stop) that varied in voice onset time, as well as F1 and F2 trajectory. In the bimodal condition infants heard a greater frequency of tokens at either end of the continuum, whereas in the unimodal condition the majority of the tokens were selected from the middle of the continuum (see Figure 4.2). This manipulation effectively simulated experience with languages that have two phoneme categories along a continuum (i.e., bimodal) versus languages that have a single phoneme category (i.e., unimodal). Following familiarization, only infants in the bimodal condition were able to discriminate the two endpoint tokens. This pattern of results suggests that experience with bimodal distributions may lead infants to form separate phoneme categories, whereas experience with unimodal distributions may encourage infants to collapse acoustic variation into a single category. In a conceptual replication of this work, Teinonen and colleagues (Teinonen, Aslin, Alku, & Csibra, 2008) found that visual speech information can also alter how infants assign distributional information to phonetic categories.

Across the first year, in addition to discovering the relevant sound categories of their native language, infants also become better at discriminating native-language phonemes (Kuhl et al., 2006; Narayan, Werker, & Beddor, 2010). In follow-up work, Maye, Weiss, and Aslin (2008) tested the prediction that distributional information also enhances phoneme discrimination, especially for contrasts that are initially difficult to discriminate. Again, infants were familiarized with either a unimodal or bimodal distribution and were tested on tokens from the endpoint of that distribution (i.e., /da/ and /ta/) or on a different contrast that varied along the same acoustic dimension (i.e., voice onset time, /ka/ and /ga/). Discrimination was only facilitated by experience with the bimodal distribution, for both the trained distribution and the novel contrast. These results suggest that infants can use distributional information to extract a specific

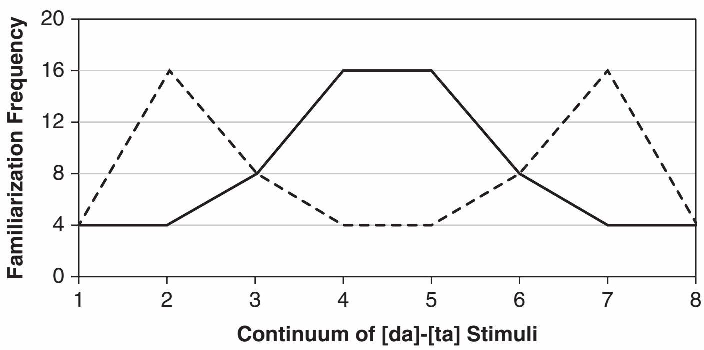

Figure 4.2 Bimodal (dashed lines) and unimodal (solid line) sound distributions used in Maye, Werker, & Gerken, 2002.

acoustic feature (e.g., voice onset time), and generalize this knowledge to novel sound distinctions. Clearly, distributional information in natural language is much more complex than the distributions presented in the lab. However, if infants are also sensitive to the distribution of speech sounds found in natural language, this sensitivity may support phonological development.

Further support for the idea that infants learn their phoneme categories through a process of distributional learning comes from work by Anderson, Morgan, and White (2003) showing that infants home in more quickly on consonant contrasts that occur more frequently in their input than on those that are heard more seldom. Consistent with this, infants perceptually narrow in on native-language vowel categories more quickly than consonant categories (Polka & Werker, 1994). Indeed, languages tend to have fewer vowels than consonants, and thus individual vowels typically occur more frequently in the input. Infants being exposed to more than one language have the added challenge of differentiating distributional patterns both within and across their languages (see Chapter 16 for further information about bilingual language acquisition). These findings support the idea that the sounds that are heard more frequently are learned more quickly, suggesting that infants are also attentive to distributional properties in natural language input.

Computational modeling work has aimed to further uncover how sensitivity to distributional statistics interacts with other learning processes to give rise to phoneme learning. Recent computational models have suggested that the competition between representations of similar speech sounds may be one way that distributional statistics are categorized into phonemes. For example, in modeling biologically plausible distributional learning for vowels Vallabha and colleagues (Vallabha, McClelland, Pons, Werker, & Amano, 2007) found that phoneme clusters emerge from experience with the environment through competitive Hebbian learning processes that strengthen often heard phonemes into clusters, and prune the connections to less often heard sounds. Taking a different approach, McMurray, Aslin, and Toscano (2009) modeled the developmental trajectory of phoneme learning and also suggest that competitive processes are vital to phoneme categorization. According to their model, early in development, phoneme categories are represented continuously, with little space between representations. However, with exposure to distributional statistics, and the addition of a winner-take-all competition mechanism, phoneme categories became sparse and more discrete across developmental time, mimicking distributional phoneme category learning. Together, these modeling studies suggest that distributional statistics interact with perceptual and neural competition mechanisms to drive the phoneme learning process.

Some research suggests that distributional learning abilities may begin to decrease around 10 months of age, when infants appear to pick up on distributional patterns only if they engage in sustained attention during familiarization (Yoshida, Pons, Maye, & Werker, 2010). Further, beyond 10 months listeners may require additional experience with distributional patterns to show learning (Yoshida et al., 2010; Maye & Gerken, 2000). This timeframe appears to be consistent with the trajectory of perceptual narrowing in infancy (Maurer & Werker, 2014) and may also coincide with infants' ability to begin to take advantage of word-level information, such as lexical items that contrast in a single vowel (e.g., *dog* vs. *dig*) or consonant (e.g., *cat* vs. *hat*) (i.e., minimal pair words) when learning overlapping phonetic categories (Feldman, Myers, White, Griffiths, & Morgan, 2013).

Nevertheless, being good at tracking distributional properties in natural language early in infancy may produce a cascading effect on subsequent language learning. Indeed, Kuhl and colleagues (Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005; Tsao, Liu, & Kuhl, 2004) provide evidence that individual differences in how quickly infants home in on their native-language sound categories are related to later language outcomes. For example, Tsao, Liu, and Kuhl (2004) found that infants' ability to discriminate vowels at 6 months predicted their productive vocabulary size at 13, 16, and 24 months. In a subsequent, larger-scale study, Kuhl and colleagues (2005) found that 7-month-olds' ability to discriminate native and non-native contrasts was negatively correlated, such that infants who are better at discriminating phonemes in their native language are worse at discriminating non-native contrasts. Additionally, performance on native and non-native contrasts predicted later language outcomes in opposite directions. Infants who were better at native-language phoneme discrimination showed larger productive vocabulary sizes at 18 and 24 months, longer mean length of utterance, and greater sentence complexity compared to the infants who were less good at discriminating native-language contrasts. Conversely, infants who showed better performance on non-native contrasts at 7 months had smaller productive vocabularies, smaller mean length of utterances, and less complex sentences when tested at 24 months.

Individual differences in speech perception skills also appear to be tied to language outcomes in atypical populations. For example, infants at familial risk for developing dyslexia show poorer phoneme discrimination at 6 months than age-matched controls (Lyytinen et al., 2001). Additionally, when tested 3 years later they produced shorter sentences and had smaller vocabularies. Longitudinal data from Boets and colleagues (Boets, Vandermosten, Poelmans, Luts, Wouters, & Ghesquière, 2011) also supports the link between speech perception skills and dyslexia; children who performed worse on a phoneme perception task in kindergarten or first grade were more likely to be diagnosed with dyslexia 2 to 3 years later. Taken together, these results suggest that early speech perception skills may influence later language outcomes in children with dyslexia (Banai & Ahissar, 2017; Boets et al., 2011; Lyytinen et al., 2001). Section 3 of this volume will provide a more in-depth review of language development in atypical populations.

Findings on individual differences from both typical and atypical populations do not directly test the relationship between distributional learning and subsequent language outcomes. However, they do suggest that infants who learn to differentiate meaningful from irrelevant sound variants at a young age continue to be more advanced language users in the second year of life and beyond. If distributional learning drives perceptual narrowing, it stands to reason that infants who are better distributional learners may ultimately be more skilled language users.

Finding words in speech

A second, and likely concurrent, challenge infants face when learning language is finding words in continuous speech. Word boundaries in spoken language are not marked by silent pauses (Cole & Jakimik, 1980), and less than 10% of the words infants hear occur in isolation (Brent & Siskind, 2001). Yet, by their first birthday, infants can understand many words and are beginning to produce their first words. Although natural languages contain various cues to word boundaries, such as prosodic patterns, phonotactic regularities, and allophonic variations, these cues are, by and large, language-specific. Thus, to use these cues infants need to know how they pattern in words in their native language. How do infants segment the speech before they know much about the patterns in their native language?

One of the most prominent hypotheses is that infants may begin to discover word boundaries, and thus segment words from fluent speech, by tracking the predictability between syllables within words and across word boundaries (e.g., Aslin, Saffran, & Newport, 1998; Saffran et al., 1996). In natural language, syllables within words tend to be highly predictive, whereas across word boundaries syllables are less predictable because words can be combined in many different ways. This type of statistical regularity is referred to as forward transitional probability (TP) when the first syllable of a word predicts the occurrence of the next syllable.

$$\text{Forward TP} = P\left(Y \middle| X\right) = \frac{frequency\left(XY\right)}{frequency\left(X\right)}$$

$$\text{Backward TP} = P\left(X \middle| Y\right) = \frac{frequency\left(XY\right)}{frequency\left(Y\right)}$$

When a given syllable predicts the occurrence of a former syllable, the statistics is referred to as backward TP and may be especially informative at the phrasal level in some languages (e.g., in the phrase *the dog* hearing *dog* predicts the occurrence of *the*, but not vice versa). Both forward and backward TPs are types of sequential co-occurrence statistics, because the predictive elements occur sequentially. Mutual information is the extent to which syllables within words are informative about each other (Swingley, 1999; 2005), and thus is essentially a statistic that incorporates predictive information in both the forward and backward directions. Finally, syllables and words also differ in their base-rate frequencies (i.e., how many times they occur; Swingley, 2005). Corpus analyses reveal that all of these sources of information are available in the input and may provide a fairly reliable language-general cue to word boundaries (Swingley, 1999).

As mentioned in the introduction, seminal work by Saffran, Aslin, and Newport (1996) demonstrated that infants are actually quite good at tracking TP information to find word boundaries in artificial language materials, suggesting that infants may begin to break into the speech stream by tracking sequential statistics in their input. This finding has been replicated numerous times (for a review, see Krogh, Vlach, & Johnson, 2013; Romberg & Saffran, 2010). Artificial languages are useful for designing tightly controlled studies and establishing infants' computational abilities, but lack the complexity associated with natural language input. In an attempt to increase ecological validity, Pelucchi, Hay, and Saffran (2009a) created familiarization corpora where we manipulated the statistics within words in a natural language. The corpus comprised naturally produced, grammatically correct, and semantically meaningful Italian sentences. Importantly, four key target words were incorporated throughout the corpus. Two of the target words had a high TP (HTP; TP = 1.0) because the syllables that made up the words never occurred anywhere else in the corpus. The other two target words had a low TP (LTP; TP = .33) because their first syllable occurred in many other words throughout the corpus (thus, TP was low in the forward direction). All four of the words were phonotactically legal in both English and Italian, and had a strong-weak (i.e., trochaic) stress pattern characteristic of disyllabic words in both languages. Using procedures very similar to those of Saffran and colleagues (1996), we found that following just 2 minutes of familiarization, 8-month-old infants preferred listening to the HTP words, even though both HTP and LTP words had been heard an equal number of times (i.e., 18 times) in the speech stream. This work suggests that infants are able to track TP information even in complex natural speech and lends support for the idea that statistical learning may play a role in early speech segmentation. In subsequent work, we have found that not only are young infants sensitive to forward TP information, but they can also use backward TP to find words in naturally spoken Italian (Pelucchi et al., 2009b).

As infants gain experience with the prosodic and phonotactic patterns of words in their native language, they may come to rely less on sequential statistics to find words in speech, in favor of language-specific cues. And indeed, statistical learning may be one avenue through which infants learn these language-specific patterns (Thiessen & Saffran, 2003). For example, using artificial language materials, Thiessen and Saffran (2003; see also Thiessen & Saffran, 2007; Johnson & Jusczyk, 2001) found that earlier in development (~ 7 months), infants rely more heavily on TP information than on stress pattern cues for identifying word boundaries.

However, by 9 months infants appear to weigh syllable stress more heavily than TP information. Thus, as infants gain more experience with their native language, their ability to track statistical patterns in speech interacts with what they know about cues to word boundaries in their native language (see also Johnson & Seidl, 2009). Fortunately, language-general and language-specific cues to word boundaries tend to converge in natural speech, and infants appear to benefit from the combined cues (Christiansen, Allen, & Seidenberg, 1998; Romberg & Saffran, 2010). Indeed, in our tasks (see Pelucchi et al., 2009a, b), infants were presented with at least two cues to word boundaries – TP information and lexical stress – and these converging cues may have made it easier for infants to track TP in the complex natural language input. In our lab, we are currently testing whether infants can track TP information in natural language when the target words have a less familiar, iambic (weak-strong) stress pattern.

Statistical-sequential learning may also be supported by other cues present in natural speech. For example, although isolated words do not make up a large portion of input to infants (Brent & Siskind, 2001), they appear to benefit from isolated words when computing TP (Lew-Williams, Pelucchi, & Saffran, 2011). Isolated words also appear to selectively support long-term memory for statistically defined words recently segmented from fluent speech by 8-month-old infants (Karaman & Hay, 2018). Work with adults further suggests that when words' base-rate frequencies are scaled up to those found in natural speech, signature patterns of statistical learning are replicated and persist in long-term memory (Frank, Tenenbaum, & Gibson, 2013). Finally, 8-month-olds can combine known words and TP information to enhance their word segmentation (Mersad & Nazzi, 2012). Similar to our discussion of the role of distributional learning in phoneme acquisition, the extant data does not provide definitive evidence infants actually use their ability to track TP information in the service of learning their native language. However, demonstrations that infants can track TP over input that increasingly approximates natural language input suggests that statistical-sequential learning may indeed be related to speech segmentation.

The link between statistical-sequential learning and language acquisition is further supported by recent research by Jill Lany and colleagues (Lany, Shoaib, Thompson, & Graf Estes, 2018) and by work with atypical populations (for a meta-analysis on the relation between statistical learning and specific language impairment, see Lammertink, Boersma, Wijnen, & Rispens, 2017). Lany and colleagues (2018) measured 15-month-old infants' performance on both a standard speech segmentation task that requires infants to track TP information in fluent speech and a task that measures their lexical knowledge, with the prediction that the infants who are better at processing native-language input may also be better statistical learners. They found that infants who were faster to look at a familiar object after it was labeled also demonstrated a preference for high TP words following familiarization with an artificial language. These findings suggest an intimate relationship between infants' sensitivity to statistical cues in the input and how well they know words in their native language.

Research with atypical populations can also reveal the potential relevance of statistical-sequential learning for language acquisition. Interestingly, 8- to 20-month-olds diagnosed with Williams Syndrome, a genetic disorder that impairs cognitive functioning but leaves language functioning relatively intact, show no impairment in tracking sequential statistics in artificial language materials (Cashon, Ha, Graf-Estes, Saffran, & Mervis, 2016). Conversely, children with Developmental Language Disorder (DLD; Bishop, Snowling, Thompson, & Greenhalgh, 2017) (sometimes also referred to as Specific Language Impairment; SLI), who demonstrate relatively intact cognition but impaired language, have been shown to have difficulty tracking sequential statistics in similar types of studies (Evans, Saffran, & Robe-Torres, 2009; see also a recent meta-analysis by Lammertink et al., 2017). Children with autism spectrum disorder (ASD), who tend to show a great deal of variability in language skills, appear to be sensitive to sequential statistics in artificial language materials (Mayo & Eigsti, 2012), but not in natural language stimuli (Kovelman et al., 2015). Kovelman and colleagues (2015) suggest that the reduced sensitivity to rhythmic information in continuous speech seen in children with Autism may impair their ability to track TP information in natural language. Further studies that explore the relationship between individual difference in language skills within atypical populations and statistical-sequential learning abilities may bolster support for the importance of statistical learning in language development.

Mapping words to meaning

A third challenge infants face when learning language is discovering how words map to meaning. Infants begin to associate words and objects early in development, and know the names of some concrete nouns as early as 6 months of age (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 2012). However, there remain challenges in determining both what sound combinations form good labels, and which of the many possible referents in the environment a given label refers to. In this way, the learning challenge is often ambiguous on multiple levels.

To overcome the challenge of discovering potential word candidates, infants can rely on what they already know about sounds. For instance, nonsense words with features consistent with those present in the native language (e.g., high phonotactic probability, high neighborhood density, stress pattern) are treated as better candidate labels than those that do not follow such patterns (e.g., Gonzalez-Gomez, Poltrock, & Nazzi, 2013; Graf Estes & Bowen, 2013; MacKenzie, Curtin, & Graham, 2012; Storkel, Bontempo, Aschenbrenner, Maekawa, & Lee, 2013).

There is also evidence that infants treat recently segmented sounds as good word candidates. For example, Graf Estes, Evans, Alibali, and Saffran (2007) found that 17-month-olds readily learn to map words segmented from an artificial language (TP = 1.0) to visual referents, but fail to map nonwords and part-words (TPs of 0.33 and 0, respectively). Moreover, we found a similar pattern of results with 17-month-olds when they were familiarized with a corpus of naturally produced Italian (Hay, Pelucchi, Graf Estes, & Saffran, 2011) – infants mapped high TP (HTP, TP = 1.0) words to novel objects, but failed to map low TP (LTP) words (forward and backward TP = .33).

In recent work from our lab, we are finding that when words are presented in carrier phrases (e.g., "Look at the casa"), older infants aged 22 to 24 months, map both HTP and LTP words to referents, but do so for different reasons. Like younger infants, older infants appear to map HTP words because of their strong internal sequential statistics. However, older infants appear to be mapping the LTP words because the syllables that compose those words are highly frequent in the corpus.[1] Thus, for older infants both sequential statistics (i.e., TP information) and syllable frequency statistics appear to feed into subsequent word learning. Interestingly, individual-differences data exploring the relationship between performance on our word learning task and vocabulary size support our findings that HTP and LTP words are mapped for different reasons; there were no significant correlations between vocabulary size and word learning for the HTP words, but there was a significant negative correlation between vocabulary size and word learning for LTP words.

In a more pared-down task that provided minimal referential support, with slightly younger 20-month-old infants, Shoaib, Wang, Hay, and Lany (2018) found that although, overall, infants did not show learning of either HTP or LTP words, performance was also correlated with their vocabulary size. Only infants with a smaller vocabulary successfully mapped the HTP words to meaning, suggesting that as infants gain native-language knowledge they become less open to treating non-typical words as good object labels. Further, like in the work from our lab described above, infants with a larger vocabulary showed a tendency for more success in mapping LTP words to referents. Together, these studies not only suggest that are HTP words treated as candidate object labels, but highlight how taking an individual-differences approach to studying statistical learning may reveal differences in underlying processes (for additional discussion of the relationship between individual differences in statistical word learning and vocabulary development, see also Lany & Saffran, 2010; 2011; Lany, 2014).

Although Mainela-Arnold and Evans (2014) have demonstrated that the ability to track TP information is related to phonological (but not semantic) processing in both children with and without DLD, to our knowledge there is only one study that has investigated how statistical-sequential learning feeds into subsequent word learning with atypical populations. Haebig, Saffran, and Weismer (2017) tested school-aged children (8 to 12 years old) with typical development, ASD, and DLD on their ability to segment words from an artificial language, map words to referents without any pre-familiarization, and a combination of the word segmentation and word mapping tasks. Although children with DLD appeared to perform at chance on the word segmentation task, they benefited most from experience with the sequential statistics in the combination task. Children diagnosed with ASD who did not have any language impairments performed similarly to typically developing children on all tasks. However, ASD children with language impairment appeared to struggle with the word mapping task, but only if they did not get prior experience with the words in the artificial language. Contrary to expectations in the combination task, across all populations, children did not map statistically defined words (i.e., words with strong internal co-occurrence statistics) to referents more easily than nonwords (made of non-adjacent syllables from the corpus). One possible interpretation of these findings, based on recent data from our lab (see above), is that mapping of the nonwords may be driven by sensitivity to syllable frequency information in the corpus, although this interpretation is speculative.

The use of these types of combined word segmentation plus word mapping tasks raises the question of whether word learning in non-ambiguous situations only unfolds in a two-stage process, where sounds are first segmented (creating a meaningless proto-lexicon) and only then mapped to referents, or if the two processes can occur simultaneously. Cunillera, Laine, Càmara, and Rodríguez-Fornells (2010) showed that adults can segment words from continuous speech and simultaneously map them to objects (see also François, Cunillera, Garcia, Laine, & Rodriguez-Fornells, 2017). Further, using a modified statistical learning task, where infants were presented with statistically defined words (TP = 1.0) in short phrases, Shukla, White, and Aslin (2011) showed that infants as young as 6 months can simultaneously extract words (TP = 1.0) and map them to visual referents when those words were aligned with prosodic cues, but fail to do so when statistics and prosody are misaligned. The authors argue that young infants may combine both statistical cues (e.g., TP) and native-language perceptual cues (e.g., prosody) to overcome simultaneous segmentation and mapping challenges.

The uncertainty with which labels and referents co-occur in natural environments poses a unique challenge for word learning (Quine, 1960). Uncertainty can take many forms. For example, the likelihood that labels and referents will co-occur can vary, and infants appear to be sensitive to these co-occurrence statistics. For example, Vouloumanos and Werker (2009) found that 18-month-olds succeeded at learning label-referent pairs when labels were consistently applied to a single referent (100% of the time) and when labels were applied to a given referent the majority of the time (i.e., 80%). Infants failed to map label-referent pairs when the co-occurring probability was much lower (i.e., 20%). These results suggest that the strength of co-occurrence between labels and referents may be an important statistical cue for word learning.

In natural learning environments, the presence of multiple labels and referents creates another source of uncertainty. A number of studies over the past decade have investigated how statistical cues may help children and adults overcome this challenge (for a review see Yu & Smith, 2012a). For instance, Smith and Yu (2008) presented 12- and 14-month-old infants with ambiguous naming events. On any given trial infants saw two referents and heard two labels; however, the mapping between individual labels and referents was not transparent. Infants could only discover label-referent pairs by comparing co-occurrences across trials (see Figure 4.3), a phenomenon known as cross-situational statistical learning, and both age groups succeeded in the task. Furthermore, even when labels contain phonetic overlap (e.g., *bon-ton*, *deet-dit*) – a common characteristic of words found in natural speech – infants at 12, 15, 17, and 20 months of age

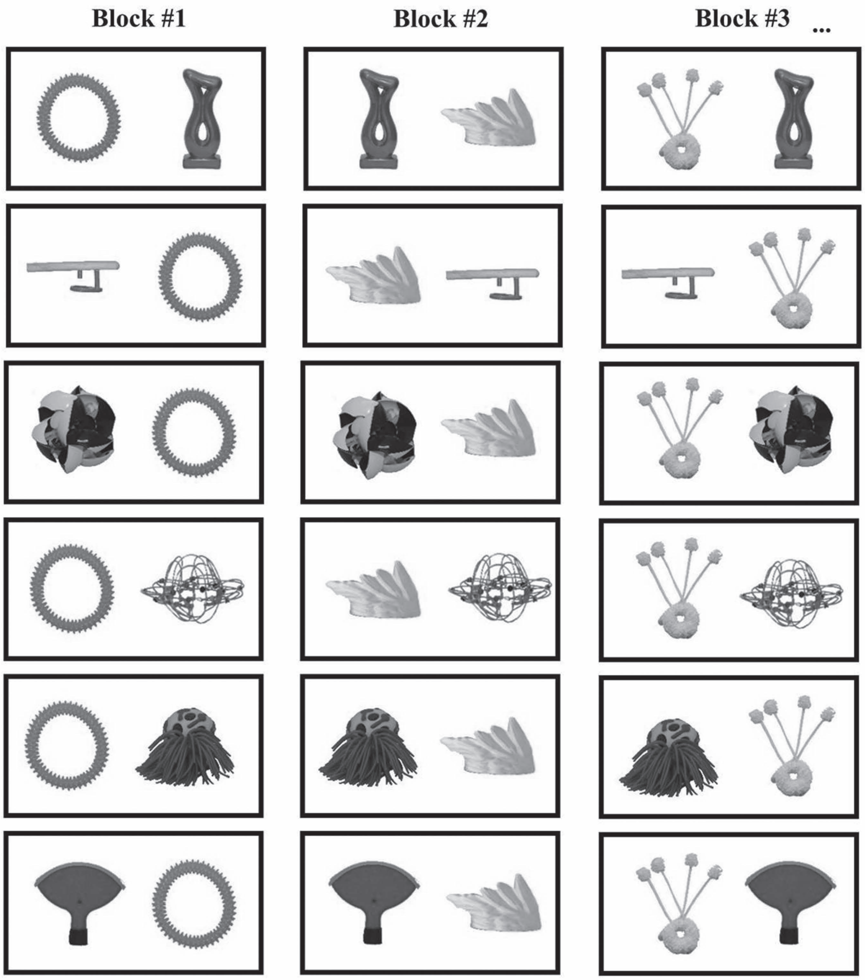

Figure 4.3 Two trials from a cross-situational learning task. Dashed lines indicate possible relations between labels and referents. For instance, by tracking label-referent co-occurrence across trials infants could map /manu/ to the circular object.

Source: The pictures from this example come from Horst and Hout (2016), and the words come from Smith and Yu (2008).

appear to map referents to labels across trials (Escudero, Mulak, & Vlach, 2016). These basic findings have been replicated in 5- to 7-year-olds (Suanda, Mugwanya, & Namy, 2014) and with children diagnosed with ASD (McGregor, Rost, Arenas, Farris-Trimble, & Stiles, 2013).

Individual-differences research suggests that cross-situational statistical learning may be related to other measures of cognitive and language development in typically developing children. For example, Scott and Fisher (2012) found that while 30-month-olds with larger vocabularies successfully mapped both transitive and intransitive verbs to actions across situations, infants with smaller vocabularies were only able to map the intransitive verbs, suggesting that cross-situational statistical learning abilities may be related to vocabulary development. In another study, Smith and Yu (2013) used a block design to test how 12- and 14-month-olds' attentional patterns affected cross-situational statistical learning and found that while learners and non-learners did not differ in how much attention they paid during training (differing from a previous report by Yu & Smith, 2011), learners had significantly larger receptive and productive vocabularies when compared to non-learners.

Using a similar approach, Vlach and Johnson (2013) explored the role of memory in cross-situational statistical learning with 16- and 20-month-olds, by presenting half of the label-referent pairs in a consecutive manner, and the other half interleaved across blocks (see Figure 4.4). Twenty-month-olds learned both the consecutive and interleaved pairs, whereas 16-month-olds only learned pairs presented in consecutive order. Neither attentional patterns nor vocabulary size predicted learning. In a follow-up study, designed to examine the relationships among age, receptive vocabulary, and recognition memory in cross-situational statistical learning, Vlach and DeBrock (2017) tested children between 22 and 66 months of age. They found that while vocabulary size and recognition memory both predicted learning, recognition memory was a stronger predictor of performance. Child age was not predictive of learning. Together these studies suggest that cross-situational statistical learning abilities may be related to the development of both language and non-linguistic cognitive skills.

Smith, Yu, and collaborators (Pereira, Smith, & Yu, 2014; Smith, Yu, & Pereira, 2011; Smith, Yu, Yoshida, & Fausey, 2015; Yurovsky, Smith, & Yu, 2013) have proposed a new approach for framing the uncertainty problem under natural contexts. By taking infants' egocentric vision into account (using lightweight head-mounted cameras), researchers have begun to understand how

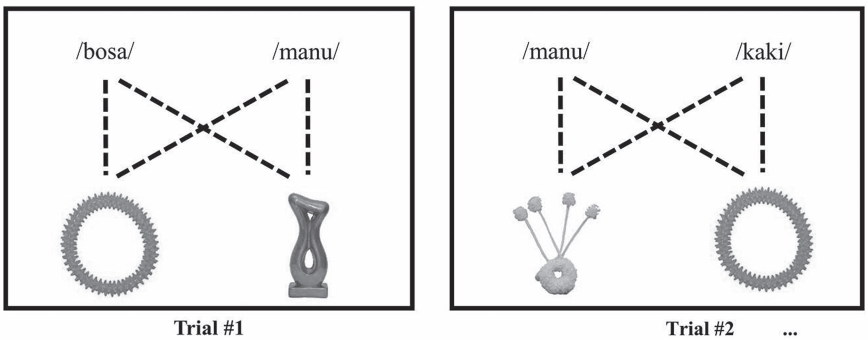

Figure 4.4 An example of the type of block design used by Vlach and Johnson (2013). On each block, one pair is always repeated consecutively across trials (massive presentation), and the other pair is presented once in each block (interleaved presentation).

Source: The pictures from this example come from Horst and Hout (2016).

cross-situational statistical learning becomes constrained by infants' own interactions with the world. For instance, 16- to 18-month-old infants disambiguate visual scenes in the laboratory by holding and looking at only one object at a time, and their parents, in turn, often name the object during these optimal learning moments, promoting word learning (Yu & Smith, 2012b). Further, even in highly cluttered natural visual and auditory learning environments, a limited set of objects fill the infant's visual field during specific activities (e.g., table, shirt, chair, and bowl during meal time) (Clerkin, Hart, Rehg, Yu, & Smith, 2017). These more constrained contexts offer ideal word learning opportunities, and indeed many of these activity-specific words are found in infants' early vocabularies (see also Bergelson & Aslin, 2017; Roy, 2009). Chapter 7 (this volume) delivers a detailed analysis of this research and its insightful advancements.

Learning rudimentary grammar

In addition to discovering sounds in language, finding words in speech, and mapping words to meaning, infants also face the challenge of learning the grammatical patterns in their native language. For example, not only do infants need to learn syntactic categories (e.g., which words are nouns, verbs, etc.), but they also need to learn how relational categories (e.g., subject, object) are ordered in their native language (e.g., is their language a subject-object-verb (SOV) language or a subject-verb-object (SVO) language?). For example, infants need to discover how past, present, and future tenses are used across non-adjacent (or long-distance) sequences (e.g., *Tomorrow* he *will* work; she *is* play*ing*), where there is an intervening element that is often irrelevant to the tense marking (e.g., *he* and *play* from the examples above). Importantly, learning grammar requires infants to be able to generalize to novel exemplars (e.g., she *is* glerp*ing*). Research has shown that infants can track multiple forms of statistical information (e.g., frequency, co-occurrence, sequential statistics, non-adjacent dependencies, etc.) and can use this information to solve many of these higher-order grammatical challenges found in language.

Early attempts to investigate the role of statistical regularities in the acquisition of rudimentary grammar were focused primarily on how infants learn word order patterns. Although in some languages word order is not particularly relevant (e.g., Turkish, Farsi), in other languages (e.g., English, Dutch) the placement of words can significantly affect the meaning of sentences (e.g., *I had my car cleaned* vs. *I had cleaned my car*). In one study, Gómez and Gerken (1999) exposed 12-month-old infants to a 2-minute familiarization with an artificial language generated by a finite-state grammar (see Figure 4.5), where some word sequences were more predictive than others. At test, not only were infants able to discriminate grammatical from ungrammatical strings based on their predictiveness, they were also able to generalize the learned grammatical patterns to novel exemplars (see Gerken, 2006, for evidence that the specific structure of the distributional information can lead to broad versus narrow generalization of grammatical patterns). Further, using a somewhat simpler design, Marcus and colleagues (Marcus, Vijayan, Rao, & Vishton, 1999) have demonstrated that infants have the ability to generalize abstract "rule-like" grammatical patterns by 7 months.

Statistical learning is a dynamic process – much as the output of word segmentation can be the input for word learning (Graf Estes et al., 2007; Hay et al., 2011), recently segmented words also feed into learning about word order. Saffran and Wilson (2003) presented 12-month-old infants with an artificial language where words (TPs within words were 1.0, while TPs across word boundaries were 0.25) were organized according to a simple finite-state grammar, and the infants were tested on their ability to discriminate novel grammatical from ungrammatical strings. In order for infants to learn which word orders were permissible in the language, they first had to track TP information to discover which sound sequences formed words. Critically,

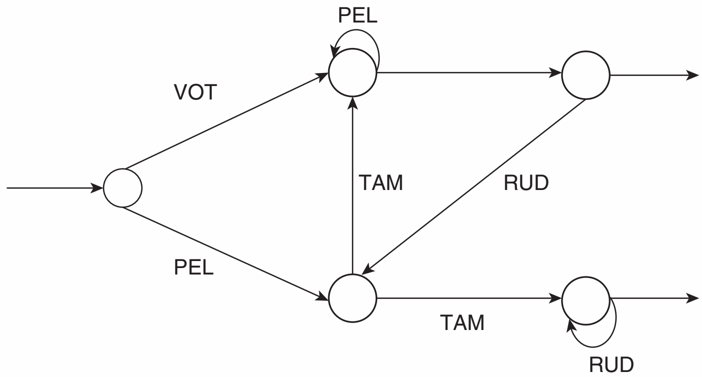

Figure 4.5 Finite-state grammar used by Gómez and Gerken (1999). Strings start on the far left and proceed through the system in the direction of the arrows (e.g., VOT PEL JIC RUD TAM RUD represents a grammatical string).

Source: Gómez and Gerken (2000).

the TP between syllables was identical in both sentence types, and so they could not be discriminated based on TP information alone. Results showed that infants were able to differentiate grammatical versus ungrammatical strings, suggesting that their ability to track TPs across syllables to find words subsequently allowed them to track word order information.

Infants' sensitivity to distributional properties in grammatical relations might also help them learn abstract syntactic categories (e.g., determiners, adjectives, nouns, verbs). For example, if English-learning children learn that determiners *the* and *a* precede nouns but not verbs, whereas auxiliaries like *was* and *is* precede verbs but not nouns, then they should be able to make inferences about the syntactic categories of novel words. Thus, learning abstract syntactic categories may play an important role in grammatical development because they can support generalization to new utterances. Corpus analyses reveal that the same syntactic categories tend to appear in highly overlapping distributional contexts, or "frequent frames" (Chemla, Mintz, Bernal, & Christophe, 2009; Mintz, 2002; 2003); however, learning those categories by tracking distributional information alone can be very difficult for infants (Gerken, Wilson, & Lewis, 2005; Gómez & LaKusta, 2004; but see Lany & Saffran, 2011).

Importantly, in natural languages, words and their syntactic categories are marked not only by distributional cues but also by phonological and semantic cues. For instance, in English, nouns tend to both have a strong-weak stress pattern and occur after determiners *the* and *a*. To investigate the role of correlated cues in the acquisition of syntactic categories, a number of studies have used an *aX bY* type of artificial language that consists of nonsense word categories *a*, *b* (e.g., similar to determiners *the* and *a*) and categories *X*, *Y* (e.g., similar to categories like nouns and verbs). In these artificial grammars, *a* category is paired with *X*, and *b* category is paired with *Y*, but not vice versa, similar to the types of co-occurrence relationships described above for English syntactic categories. For example, using an *aX bY* paradigm, Gómez and LaKusta (2004) found that 12-month-olds were able to form categories by associating the distributional information of *a* and *b* words with a particular phonological feature of *X* and *Y* words (i.e., *X* words were monosyllabic and *Y* words were disyllabic). In another study, Lany and colleagues (2017) found that at 15 months, infants' ability to track these types of correlated cues in artificial language materials is related to individual differences in how efficiently infants process familiar words in their native language.

In another series of experiments, Lany and Saffran (2010) sought to investigate whether these types of correlated cues also facilitate the grouping of words into categories and subsequent mapping of those words to different categories of referents (e.g., animals vs. vehicles). Twenty-two-month-old infants were familiarized with an *aX bY* language where either *a* words (*ong* and *erd*) predicted two syllable words (e.g., *coomo*, *loga*) and *b* words (*atl* and *ush*) predicted one syllable words (e.g., *deech*, *jic*), or a control condition where *a* and *b* words each predicted one and two syllable words. Instead of testing infants on novel grammatical versus ungrammatical strings, infants went directly into a label-referent mapping task where *aX* sequences were paired with referents from one lexical category (e.g., animals) and *bY* sequences were paired with referents from another lexical category (e.g., vehicles). Infants learned individual referents and lexical categories when distributional and phonological cues were correlated (i.e., co-occurrence and number of syllables), and successfully generalized untrained *aX* and *bY* sequences to novel exemplars of the referent categories. Infants failed to learn when the cues were not correlated during familiarization even though they received the same experience during label-referent training. In a subsequent study, Lany and Saffran (2011) found that infants with more advanced vocabulary and grammatical skills (as measured by parental report) could use distributional cues alone to identify a novel referent from the learned category (i.e., upon hearing an *a* word infants would look at the novel exemplar of an animal), but they could not use phonological cues alone (i.e., when hearing an untrained two syllable word they failed to look at the novel animal exemplar). The reverse pattern was found for infants with less advanced vocabulary and grammatical skills (Lany & Saffran, 2011; see also Lany, 2014). This set of findings suggests that infants track correlations among distributional, phonological, and semantic information when learning words. They also indicate that these cues might be weighted differently depending on individual differences in infants' vocabulary and grammatical development.

Many of the studies that we have reviewed thus far have focused on infants' and children's ability to track adjacent dependencies (e.g., units that co-occur without any intervening information). However, natural languages also contain many sequential but non-adjacent dependencies, where the intervening information between co-occurring elements can vary (e.g., *is* writ*ing*, *is* read*ing*, *is* sing*ing*). Infants appear to be able to learn these types of morphosyntactic non-adjacent dependencies if there is sufficient variability in the intervening element (e.g., $AX_1B, AX_2B, \ldots, AX_{24}B$), but fail to learn when the intervening element only takes on a limited number of forms (Gómez, 2002; but see Vuong, Meyer, & Christinsen, 2016, for conflicting findings with adults). When variability in the intervening elements is low, learners may track individual items (tokens) in the input, but may fail to appreciate subtler patterns that occur between item types (e.g., the grammar). Conversely, Gómez (2002) suggests that high variability in the intervening elements decreases the co-occurrence statistics between adjacent elements, causing infants to redirect their attention from adjacent to more informative non-adjacent dependencies. Conceptual replications of this work, with both artificial and natural languages, have found that infants as young as 15 months appear to be sensitive to non-adjacent dependencies (e.g., Gómez & Maye, 2005; Lany & Gómez, 2008; van Heugten & Johnson, 2010). Further, although 12-month-olds appear to struggle learning non-adjacent dependencies, they appear to be able to use the adjacent relations to learn non-adjacent ones (Lany & Gómez, 2008).

Some languages also contain phonological non-adjacent dependencies. Unlike in English, in Turkish, vowels within a word tend to agree in their place of articulation, a feature called vowel harmony. For instance, the word for child (*çocuk*) is legal in Turkish because the /o/ and /u/ are both back vowels. Similarly, *kedi* (cat) contains all front vowels. Except for in borrowed words (e.g., *selam*, "hello", borrowed from Arabic), front and back vowels do not typically occur in the same words. A recent study by Mintz, Walker, Welday, and Kidd (2018) found that 7-month-old

English-learning infants can use vowel harmonic non-adjacent dependencies to segment words from fluent speech, suggesting that even without exposure to vowel harmony in their native language, infants are sensitive to these types of phonological non-adjacent dependencies early in life.

Although there are a number of studies that have explored individual differences in learning rudimentary grammar, including non-adjacent dependencies, and grammatical proficiency in adolescents and adults (e.g., Erickson, Kaschak, Thiessen, & Berry, 2016; Hsu, Tomblin, & Christiansen, 2014; Misyak & Christiansen, 2010; Misyak, Christiansen, & Tomblin, 2010; Plante, Gómez, & Gerken, 2002), to our knowledge, these relationships have not been as extensively explored in infancy and early childhood. On reason for the lack of focus on individual differences is because developmental studies on statistical learning have tended to use tasks that measure group level effects, where the aim has been to demonstrate that children of a certain age are able to learn certain types of statistical patterns. As interests have moved beyond proof of ability demonstrations, to exploring how individual differences in statistical learning map onto other types of perceptual and/or cognitive functioning, additional methodological considerations are necessary (e.g., see Siegelman, Bogaerts, & Frost, 2017, for pitfalls and solutions in studying individual differences in statistical learning). Certainly, the work by Lany and colleagues (Lany, 2014; Lany & Saffran, 2010, 2011; Lany et al., 2017; Shoaib et al., 2018) provides an exception to this focus on group level effects (see above). Further, while some recent individual-differences work by Kidd and Arciuli (2016) has found that children's ability to track statistics in a visual task is related to their grammatical knowledge (see also work by Conway, Pisoni, Anaya, Karpicke, and Hennqing (2011) on children with cochlear implants), much less is known about how tracking statistics in linguistic input is related to grammatical development. There have, however, been a few studies looking at individual differences in non-adjacent dependency learning in atypical populations. For example, a recent study by Iao, Ng, Wong, and Lee (2017) demonstrated that Cantonese-speaking children with DLD show difficulties in learning non-adjacent dependencies. Similarly, Kerkhoff and colleagues (2013) found that 18-month-olds at familial risk for dyslexia have deficits in non-adjacent dependency learning, perhaps accounting for the characteristic delays in grammatical and phonological processing seen in children with dyslexia. Although these studies have not explicitly explored the role of working memory capacity in the learning of non-adjacent dependencies, the two are likely related (Santelmann & Jusczyk, 1998; but see Hsu et al., 2014). Further research on individual differences in rudimentary grammar learning in infancy and early childhood may strengthen our understanding of the variability seen in early grammatical development.

Conclusions

In this chapter, we have reviewed some of the extant evidence that has accumulated, especially over the past 20 years, suggesting that infants are equipped with sophisticated computational abilities and can use them to learn many of the types of structures found in natural languages. Although the studies that we have presented have focused exclusively on infants' ability to track statistics in linguistic input, it is important to note that these abilities are not specific to either humans or speech. For example, evidence for distributional learning has been observed in other acoustic domains (e.g., music: Ong, Burnham, & Stevens, 2017), sensory modalities (e.g., vision: Duffy, Huttenlocher, & Crawford, 2006), and taxa (e.g., rats: Pons, 2006). Statistical-sequential learning has been demonstrated with visual (Fiser & Aslin, 2001; Kirkham, Slemmer, & Johnson, 2002), tactile (Conway & Christiansen, 2005), and non-speech auditory stimuli (e.g., Hay & Saffran, 2012; Saffran, 2003; Saffran, Johnson, Aslin, & Newport, 1999), and in a number of species (rats: Toro & Trobalón, 2005; cotton-top tamarins: Hauser, Newport, & Aslin, 2001; and songbirds: Chen & ten Cate, 2015; Takashi, Yamada, & Okanoya, 2010). Finally, a number of different species demonstrate evidence of being able to learn non-adjacent dependency relations (songbirds: Chen & ten Cate, 2017; primates: Newport, Hauser, Spaepen, & Aslin, 2004; Sonnweber, Ravignani, & Fitch, 2015) and rudimentary grammar (cotton-top tamarins: Fitch & Hauser, 2004; Saffran, Hauser, Seibel, Kapfhamer, Tsao, & Cushman, 2008; songbirds: Gentner, Fenn, Morgaliash, & Nusbaum, 2006), although not all non-human animals appear to be able to generalize grammatical patterns to new exemplars (for a review see ten Cate & Okanoya, 2012). The degree of commonality underlying all of these findings remains unknown; however, modeling (e.g., Perruchet & Vintner, 1998; Thiessen & Pavlik, 2013; Thiessen, 2017) and neuroimaging studies (for a review, see Schapiro & Turk-Browne, 2015) are beginning to answer some of these questions.

Although we have provided ample evidence that infants can track statistical regularities in the lab, we would be remiss to ignore that null findings are also prevalent in the field. For example, within the segmentation literature, infants appear to have a difficult time tracking statistics across words with variable length (e.g., Johnson & Tyler, 2010, but see Mersad & Nazzi, 2012), or if they have the expectation that words are one length (e.g., two syllables) and then the words in the speech stream are another length (e.g., three syllables; Lew-Williams & Saffran, 2012). Further, according to two recent meta-analyses (Cristia, 2018; see also Lewis et al., 2018), the evidence for distribution learning of phoneme categories has not proven to be very robust, partially because a number of researchers have failed to replicate the original Maye, Werker, and Gerken (2002) study, and of the studies that have replicated it, many have had relatively small effect sizes.

Null effects and failure to replicate are perhaps not particularly surprising given that infant methodologies are sometimes fragile, as infants are typically provided with very short periods of exposure (e.g., often ~ 2 to 10 minutes) to the statistical structures they are subsequently tested on, only a subset of infants may show learning in a given study, and small sample sizes limit statistical power. Indeed, in one recent meta-analysis by Black and Bergmann (2017), incorporating 68 experiments, and data from 1,454 infants, there was an overall significant but small effect size across studies examining infants' ability to track TP information in continuous speech. These results suggest that although it is highly likely that infants do possess these computational abilities, there is a need for increasing the power and robustness of statistical learning studies.

Some of these replication issues may also derive from the fact that much of the developmental work on statistical learning has relied exclusively on offline measures, which typically measure novelty versus familiarity preferences based on looking time differences during a post-familiarization phase. Kuppuraj and colleagues (Kuppuraj, Duta, Thompson, & Bishop, 2018) argue that these offline measures may confound effects of encoding and memory, use testing procedures that may interfere with the statistics of the initial encoding, and fail to capture how statistical regularities are tracked over time. Although online measures have been successfully and reliably used to study statistical learning in adults (e.g., Batterink & Paller, 2017; Kuppuraj et al., 2018; Misyak, Christiansen, & Tomblin, 2010), many of the common forms of infant methodologies do not lend themselves well to collecting online measures (for an exception see Romberg & Saffran, 2013). Finally, large-scale, multi-lab replication efforts, such as those outlined in the various ManyBabies projects (e.g., Frank et al., 2017), will ultimately help researchers to better understand the robustness of a variety of developmental phenomena, including statistical learning.

Although laboratory studies have revealed a great deal about infants' computational abilities, they do not demonstrate that infants actually use these abilities in the service of learning language. And indeed, given currently available methodology, proving this relationship will likely be quite difficult. Throughout this chapter, however, we have reviewed studies exploring how individual differences in statistical learning might be related to individual differences in language proficiency, in both typical and atypical populations. While most of this work does not provide direct evidence that statistical learning drives language acquisition, and indeed we would never claim that statistical learning is the only avenue through which infants learn language, it does support the hypothesis that statistical learning and language learning may be related. Further, dissociations across different atypical populations (e.g., relatively intact language and statistical learning abilities in infants with Williams Syndrome (Cashon et al., 2016), but relatively impaired language and statistical learning in infants with DLD (Evans et al., 2009)) also support the hypothesis that statistical learning and language acquisition may be interconnected.

Given the evidence for a connection between statistical learning and language learning it is not surprising that there are a number of recent recommendations for interventions for clinical populations that build on statistical learning approaches. While some interventions have relied solely on graded approaches, where training begins with simplified input that gradually increases in complexity, those built on statistical learning approaches have emphasized the importance of implicitly providing patients with the types of variability and complexity found in natural language input from the outset (e.g., Alt, Meyers, & Ancharski, 2012). Consistent with this idea, Plante and colleagues (Plante et al., 2014) found that the variability of exemplars used by clinicians, rather than the number of repetitions of a single token, helped children with DLD learn grammatical morphemes (e.g., plural marker *s*, past tense *ed*) in a conversational recasting task (see also Leonard & Deevy, 2017). Researchers and clinicians should continue to consider interventions aimed at helping children extract statistical patterns in the input as a potential first step in improving language outcomes in vulnerable populations.

In addition to individual-differences research, another approach to addressing the question of whether and how infants use their computational abilities for language learning outside the laboratory involves presenting infants with more naturalistic learning challenges. In our lab, we have taken a number of approaches to increase the ecological validity of our measures. First, we tend to study statistical learning using complex natural language materials (i.e., Italian: Karaman & Hay, 2018; Hay et al., 2011; Pelucchi et al., 2009a, 2009b). Further, we have explored how statistical learning feeds into subsequent word learning (e.g., Hay et al., 2011; Shoaib et al., 2018) and how statistical learning might support long-term memory in infants (e.g., Karaman & Hay, 2018). Finally, we are currently examining both the robustness and specificity of statistical learning, by testing infants in more naturalistic/noisy listening environments. As we are able to demonstrate that infants' computational abilities stand up to the challenges inherent to language learning outside of the lab, we will be able to further elucidate the potential role for statistical learning in language development.

Note

1 In order to lower the TP of the LTP words in both the forward and the backward direction, both the first and second syllables of the LTP words were presented throughout the corpus. Thus, while infants heard the syllables of the HTP words a total of 18 times, they heard the syllables of LTP words a total of 54 times each – 18 times in the LTP words themselves and then an additional 36 times throughout the corpus.

References

Alt, M., Meyers, C., & Ancharski, A. (2012). Using principles of learning to inform language therapy design for children with specific language impairment. *International Journal of Language & Communication Disorders*, *47*(5), 487–498. doi:10.1111/j.1460-6984.2012.00169.x

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*(2), 239–273. doi:10.1017/S030500091400049X

Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A statistical basis for speech sound discrimination. *Language and Speech*, *46*(2–3), 155–182. doi:10.1177/00238309030460020601

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. doi:10.1016/S0010-0277(98)00075-4

Banai, K., & Ahissar, M. (2017). Poor sensitivity to sound statistics impairs the acquisition of speech categories in dyslexia. *Language, Cognition and Neuroscience*, 1–12. doi:10.1080/23273798.2017.1408851

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45. doi:10.1016/j.cortex.2017.02.004

Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences*, *114*(49), 12916–12921. doi:10.1073/pnas.1712966114

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258. doi:10.1073/pnas.1113380109

Bishop, D.V., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068–1080. doi:10.1111/jcpp.12721

Black, A., & Bergmann, C. (2017). *Quantifying infants' statistical word segmentation: A meta-analysis*. Paper presented at the Proceedings of the 39th Annual Conference of the Cognitive Science Society.

Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., & Ghesquière, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research in Developmental Disabilities*, *32*(2), 560–570. doi:10.1016/j.ridd.2010.12.020

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33–B44. doi:10.1016/S0010-0277(01)00122-6

Cashon, C. H., Ha, O-R., Estes, K. G., Saffran, J. R., & Mervis, C. B. (2016). Infants with Williams syndrome detect statistical regularities in continuous speech. *Cognition*, *154*, 165–168. doi:10.1016/j.cognition.2016.05.009

Chemla, E., Mintz, T. H., Bernal, S., & Christophe, A. (2009). Categorizing words using "frequent frames": What cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental Science*, *12*(3), 396–406. doi:10.1111/j.1467-7687.2009.00825.x

Chen, J., & ten Cate, C. (2015). Zebra finches can use positional and transitional cues to distinguish vocal element strings. *Behavioural Processes*, *117*, 29–34. doi:10.1016/j.beproc.2014.09.004

Chen, J., & ten Cate, C. (2017). Bridging the gap: Learning of acoustic nonadjacent dependencies by a songbird. *Journal of Experimental Psychology: Animal Learning and Cognition*, *43*(3), 295. doi:10.1037/xan0000145

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. doi:10.1080/016909698386528

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711). doi:10.1098/rstb.2016.0055

Cole, R., & Jakimik, J. (1980). A model of speech perception. In R. Cole (Ed.), *Perception and 734 production of fluent speech* (pp. 133–163). Hillsdale, NJ: Lawrence Erlbaum.

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. doi:10.1037/0278-7393.31.1.24

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. doi:10.1111/j.1467-7687.2010.00960.x

Cristia, A. (2018). Can infants learn phonology in the lab? A meta-analytic answer. *Cognition*, *170*, 312–327. doi:10.1016/j.cognition.2017.09.016

Cunillera, T., Laine, M., Càmara, E., & Rodríguez-Fornells, A. (2010). Bridging the gap between speech segmentation and word-to-world mappings: Evidence from an audiovisual statistical learning task. *Journal of Memory and Language*, *63*(3), 295–305. doi:10.1016/j.jml.2010.05.003

Duffy, S., Huttenlocher, J., & Crawford, L. E. (2006). Children use categories to maximize accuracy in estimation. *Developmental Science*, *9*(6), 597–603. doi:10.1111/j.1467-7687.2006.00538.x

Erickson, L., Kaschak, M., Thiessen, E., & Berry, C. (2016). Individual differences in statistical learning: Conceptual and measurement issues. *Collabra: Psychology*, *2*(1). http://dx.doi.org/10.1525/ collabra.41

Escudero, P., Mulak, K. E., & Vlach, H. A. (2016, September). Infants encode phonetic detail during cross-situational word learning. *Frontiers in Psychology*, 7, 1–11. doi:10.3389/fpsyg.2016.01419

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. doi:10.1044/1092-4388(2009/07-0189)

Feldman, N. H., Myers, E. B., White, K. S., Griffiths, T. L., & Morgan, J. L. (2013). Word-level information influences phonetic learning in adults and infants. *Cognition*, *127*(3), 427–438. doi:10.1016/j.cognition.2013.02.007

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. doi:10.1111/1467-9280.00392

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *303*(5656), 377–380. doi:10.1126/science.1089401

François, C., Cunillera, T., Garcia, E., Laine, M., & Rodriguez-Fornells, A. (2017). Neurophysiological evidence for the interplay of speech segmentation and word-referent mapping during novel word learning. *Neuropsychologia*, *98*, 56–67. doi:10.1016/j.neuropsychologia.2016.10.006

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., . . . Lew-Williams, C. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, *22*(4), 421–435. doi:10.1111/infa.12182

Frank, M. C., Tenenbaum, J. B., & Gibson, E. (2013). Learning and long-term retention of large-scale artificial languages. *PLoS One*, *8*, e52500. doi:10.1371/journal.pone.0052500

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, *440*(7088), 1204. doi:10.1038/nature04675

Gerken, L. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition*, *98*(3), B67–B74. doi:10.1016/j.cognition.2005.03.003

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, *32*(2), 249–268. doi:10.1017/S0305000904006786

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436. doi:10.1111/1467-9280.00476

Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–135. doi:10.1016/S0010-0277(99)00003-7

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*(5), 178–186. doi:10.1016/S1364-6613(00)01467-4

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. doi:10.1111/j.1467-7687.2004.00381.x

Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, 7(2), 183–206. doi:10.1207/s15327078in0702_4

Gonzalez-Gomez, N., Poltrock, S., & Nazzi, T. (2013). A "Bat" is easier to learn than a "Tab": Effects of relative phonotactic frequency on infant word learning. *PLoS One*, *8*(3). doi:10.1371/journal.pone.0059601

Graf Estes, K., & Bowen, S. (2013). Learning about sounds contributes to learning about words: Effects of prosody and phonotactics on infant word learning. *Journal of Experimental Child Psychology*, *114*(3), 405–417. doi:10.1016/j.jecp.2012.10.002

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. doi:10.1111/j.1467-9280.2007.01885.x

Haebig, E., Saffran, J. R., & Ellis Weismer, S. (2017). Statistical word learning in children with autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry*, *58*(11), 1251–1263. doi:10.1111/jcpp.12734

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53–B64. doi:10.1016/S0010-0277(00)00132-3

Hay, J. F., Pelucchi, B., Estes, K. G., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology*, *63*(2), 93–106. doi:10.1016/j.cogpsych.2011.06.002

Hay, J. F., & Saffran, J. R. (2012). Rhythmic grouping biases constrain infant statistical learning. *Infancy*, *17*(6), 610–641. doi:10.1111/j.1532-7078.2011.00110.x

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*(4), 1393–1409. doi:10.3758/s13428-015-0647-3

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*, 175. doi:10.3389/fpsyg.2014.00175

Iao, L. S., Ng, L. Y., Wong, A. M. Y., & Lee, O. T. (2017). Nonadjacent dependency learning in Cantonese-speaking children with and without a history of specific language impairment. *Journal of Speech, Language, and Hearing Research*, *60*(3), 694–700. doi:10.1044/2016_JSLHR-L-15-0232

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. doi:10.1006/jmla.2000.2755

Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science*, *12*(1), 131–141. doi:10.1111/j.1467-7687.2008.00740.x

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339–345. doi:10.1111/j.1467-7687.2009.00886.x

Karaman, F., & Hay, J. F. (2018). The longevity of statistical learning: When infant memory decays, isolated words come to the rescue. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(2), 221. doi:10.1037/xlm0000448

Kerkhoff, A., De Bree, E., De Klerk, M., & Wijnen, F. (2013). Non-adjacent dependency learning in infants at familial risk of dyslexia. *Journal of Child Language*, *40*(1), 11–28. doi:10.1017/S0305000912000098

Kidd, E., & Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development*, *87*(1), 184–193. doi:10.1111/cdev.12461

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. doi:10.1016/S0010-0277(02)00004-5

Kovelman, I., Wagley, N., Hay, J. S., Ugolini, M., Bowyer, S. M., Lajiness-O'Neill, R., & Brennan, J. (2015). Multimodal imaging of temporal processing in typical and atypical language development. *Annals of the New York Academy of Sciences*, *1337*(1), 7–15. doi:10.1111/nyas.12688

Krogh, L., Vlach, H., & Johnson, S. P. (2013). Statistical learning across development: Flexible yet constrained. *Frontiers in Psychology*, *3*, 598. doi:10.3389/fpsyg.2012.00598

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the "critical period". *Language Learning and Development*, *1*(3–4), 237–264. doi:10.1080/15475441.2005.9671948

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2). doi:10.1111/j.1467-7687.2006.00468.x

Kuhl, P. K., Tsao, F. M., Liu, H. M., Zhang, Y., & Boer, B. (2001). Language/culture/mind/brain. *Annals of the New York Academy of Sciences*, *935*(1), 136–174. doi:10.1111/j.1749-6632.2001.tb03478.x

Kuppuraj, S., Duta, M., Thompson, P., & Bishop, D. (2018). Online incidental statistical learning of audiovisual word sequences in adults: A registered report. *Royal Society Open Science*, *5*(2). doi:10.1098/rsos.171678

Lalonde, C. E., & Werker, J. F. (1995). Cognitive influences on cross-language speech perception in infancy. *Infant Behavior and Development*, *18*(4), 459–475. doi:10.1016/0163-6383(95)90035-7

Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2017). Statistical learning in specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3474–3486. doi:10.1044/2017_JSLHR-L-16-0439

Lany, J. (2014). Judging words by their covers and the company they keep: Probabilistic cues support word learning. *Child Development*, *85*(4), 1727–1739. doi:10.1111/cdev.12199

Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, *19*(12), 1247–1252. doi:10.1111/j.1467-9280.2008.02233.x

Lany, J., & Saffran, J. R. (2010). From statistics to meaning: Infants' acquisition of lexical categories. *Psychological Science*, *21*(2), 284–291. doi:10.1177/0956797609358570

Lany, J., & Saffran, J. R. (2011). Interactions between statistical and semantic information in infant language development. *Developmental Science*, *14*(5), 1207–1219. doi:10.1111/j.1467-7687.2011.01073.x

Lany, J., Shoaib, A., Thompson, A., & Estes, K. G. (2018). Infant statistical-learning ability is related to real-time language processing. *Journal of Child Language*, 1–24. doi:10.1017/S0305000917000253

Leonard, L. B., & Deevy, P. (2017). The changing view of input in the treatment of children with grammatical deficits. *American Journal of Speech-Language Pathology*, *26*(3), 1030–1041. doi:10.1044/2017_AJSLP-16-0095

Lewis, M., Braginsky, M., Tsuji, S., Bergmann, C., Piccinini, P., Cristia, A., & Frank, M. C. (2018). *A quantitative synthesis of early language acquisition using meta-analysis*. https://doi.org/10.31234/osf.io/htsjm

Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science*, *14*(6), 1323–1329. doi:10.1111/j.1467-7687.2011.01079.x
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. doi:10.1016/j.cognition.2011.10.007
Lotto, A. J., Sato, M., & Diehl, R. L. (2004). Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/. *From Sound to Sense*, *50*, C381–C386.
Lyytinen, H., Ahonen, T., Eklund, K., Guttorm, T. K., Laakso, M. L., Leinonen, S., . . . Richardson, U. (2001). Developmental pathways of children with and without familial risk for dyslexia during the first years of life. *Developmental Neuropsychology*, *20*(2), 535–554. doi:10.1207/S15326942DN2002_5
MacKenzie, H., Curtin, S., & Graham, S. A. (2012). 12-month-olds' phonotactic knowledge guides their word-object mappings. *Child Development*, *83*(4), 1129–1136. doi:10.1111/j.1467-8624.2012.01764.x
Mainela-Arnold, E., & Evans, J. L. (2014). Do statistical segmentation abilities predict lexical-phonological and lexical-semantic abilities in children with and without SLI? *Journal of Child Language*, *41*(2), 327–351. doi:10.1017/S0305000912000736
Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77–80. doi:10.1126/science.283.5398.77
Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*(2), 154–178. doi:10.1002/dev.21177
Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. *Proceedings of the 24th Annual Boston University Conference on Language Development*, *2*, 522–533.
Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*(1), 122–134. doi:10.1111/j.1467-7687.2007.00653.x
Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111. doi:10.1016/S0010-0277(01)00157-3
Mayo, J., & Eigsti, I-M. (2012). Brief report: A comparison of statistical learning in school-aged children with high functioning autism and typically developing peers. *Journal of Autism and Developmental Disorders*, *42*(11), 2476–2485. doi:10.1007/s10803-012-1493-0
McGregor, K. K., Rost, G., Arenas, R., Farris-Trimble, A., & Stiles, D. (2013). Children with ASD can use gaze in support of word recognition and learning. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*(7), 745–753. doi:10.1111/jcpp.12073
McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science*, *12*(3), 369–378. doi:10.1111/j.1467-7687.2009.00822.x
Mersad, K., & Nazzi, T. (2012). When mommy comes to the rescue of statistics: Infants combine top-down and bottom-up cues to segment speech. *Language Learning and Development*, *8*(3), 303–315. doi:10.1080/15475441.2011.609106
Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition*, *30*(5), 678–686. doi:10.3758/BF03196424
Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, *90*(1), 91–117. doi:10.1016/S0010-0277(03)00140-9
Mintz, T. H., Walker, R. L., Kidd, C., & Welday, A. (2018). Infants' sensitivity to vowel harmony and its role in segmenting speech. *Cognition*, *171*, 95–107. doi:10.1016/j.cognition.2017.10.020
Misyak, J. B., & Christiansen, M. H. (2010, January). When "more" in statistical learning means 'less' in language: Individual differences in predictive processing of adjacent dependencies. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *32*(32).
Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, *1*, 31. doi:10.3389/fpsyg.2010.00031
Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, *13*(3), 407–420. doi:10.1111/j.1467-7687.2009.00898.x
Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II: Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, *49*(2), 85–117. doi:10.1016/j.cogpsych.2003.12.002
Ong, J. H., Burnham, D., & Stevens, C. J. (2017). Learning novel musical pitch via distributional learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(1), 150. doi:10.1037/xlm0000286
Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009a). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244–247. doi:10.1016/j.cognition.2009.07.011

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009b). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–685. doi:10.1111/j.1467-8624.2009.01290.x

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin and Review*, *21*(1), 178–185. doi:10.3758/s13423-013-0466-4

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246–263. doi:10.1006/jmla.1998.2576

Plante, E., Gomez, R., & Gerken, L. (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders*, *35*(5), 453–462. doi:10.1016/S0021-9924(02)00094-1

Plante, E., Ogilvie, T., Vance, R., Aguilar, J. M., Dailey, N. S., Meyers, C., . . . Burton, R. (2014). Variability in the language input to children enhances learning in a treatment context. *American Journal of Speech-Language Pathology*, *23*(4), 530–545. doi:10.1044/2014_AJSLP-13-0038

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421–435. doi:10.1037/0096-1523.20.2.421

Pons, F. (2006). The effects of distributional learning on rats' sensitivity to phonetic information. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(1), 97–101. doi:10.1037/0097-7403.32.1.97

Quine, W. A. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906–914. doi:10.1002/wcs.78

Romberg, A. R., & Saffran, J. R. (2013). All together now: Concurrent learning of multiple structures in an artificial language. *Cognitive Science*, *37*(7), 1290–1320. doi:10.1111/cogs.12050

Roy, D. (2009, 13–20). *New horizons in the study of child language acquisition*. Proceedings of the 10th Annual Conference of the International Speech Communication Association (INTERSPEECH 2009).

Saffran, J. R. (2003). Absolute pitch in infancy and adulthood: The role of tonal structure. *Developmental Science*, *6*(1), 35–43. doi:10.1111/1467-7687.00250

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. doi:10.1126/science.274.5294.1926

Saffran, J. R., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., & Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition*, *107*(2), 479–500. doi:10.1016/j.cognition.2007.10.010

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. doi:10.1016/S0010-0277(98)00075-4

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181–203. doi:10.1146/annurev-psych-122216-011805

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*(2), 101–105. doi:10.1111/j.1467-9280.1997.tb00690.x

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy*, *4*(2), 273–284. doi:10.1207/S15327078IN0402_07

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition*, *69*(2), 105–134. doi:10.1016/S0010-0277(98)00060-2

Schapiro, A., & Turk-Browne, N. (2015). Statistical learning. *Brain Mapping: An Encyclopedic Reference*, *3*, 501–506. doi:10.1016/B978-0-12-397025-1.00276-1

Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*(2), 163–180. doi:10.1016/j.cognition.2011.10.010

Shoaib, A., Wang, T., Hay, J. F., & Lany, J. (2018). Do infants learn words from statistics? Evidence from English-learning infants hearing Italian. *Cognitive Science*, *42*, 3083–3099. doi: 10.1111/cogs.12673

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(15), 6038–6043. doi:10.1073/pnas.1017617108

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods*, *49*(2), 418–432. doi:10.3758/s13428-016-0719-z

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. doi:10.1016/j.cognition.2007.06.010

Smith, L. B., & Yu, C. (2013). Visual attention is not enough: Individual differences in statistical word-referent learning in infants. *Language Learning and Development*, *9*(1), 25–49. doi:10.1080/15475441.2012.707104

Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, *14*(1), 9–17. doi:10.1111/j.1467-7687.2009.00947.x

Smith, L. B., Yu, C., Yoshida, H., & Fausey, C. M. (2015). Contributions of head-mounted cameras to studying the visual environments of infants and young children. *Journal of Cognition and Development*, *16*(3), 407–419. doi:10.1080/15248372.2014.933430

Sonnweber, R., Ravignani, A., & Fitch, W. T. (2015). Non-adjacent visual dependency learning in chimpanzees. *Animal Cognition*, *18*(3), 733–745. doi:10.1007/s10071-015-0840-x

Storkel, H. L., Bontempo, D. E., Aschenbrenner, A. J., Maekawa, J., & Lee, S. Y. (2013). The effect of incremental changes in phonotactic probability and neighborhood density on word learning by preschool children. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1689–1700. doi:10.1044/1092-4388(2013/12-0245)

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology*, *126*, 395–411. doi:10.1016/j.jecp.2014.06.003

Swingley, D. (1999). Conditional probability and word discovery: A corpus analysis of speech to infants. In M. Hahn & S. C. Stones (Eds.), *Proceedings of the 21st annual conference of the cognitive science society* (pp. 724–729). Mahwah, NJ: Lawrence Erlbaum Associates.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*(1), 86–132. doi:10.1016/j.cogpsych.2004.06.001

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. doi:10.1016/j.cognition.2008.05.009

ten Cate, C., & Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: Natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 1984–1994. doi:10.1098/rstb.2012.0055

Thiessen, E. D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*, 1711. doi:10.1098/rstb.2016.0056

Thiessen, E. D., & Pavlik, P. I. (2013). iMinerva: A mathematical model of distributional statistical learning. *Cognitive Science*, *37*(2), 310–343. doi:10.1111/cogs.12011

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. doi:10.1037/0012-1649.39.4.706

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. doi:15475440709337001

Tincoff, R., & Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, *17*(4), 432–444. doi:10.1111/j.1532-7078.2011.00084.x

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Attention, Perception, & Psychophysics*, *67*(5), 867–875. doi:10.3758/BF03193539

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. doi:10.1111/j.1467-8624.2004.00726.x

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, *104*(33), 13273–13278. doi:10.1073/pnas.0705369104

Van Heugten, M., & Johnson, E. K. (2010). Linking infants' distributional learning abilities to natural language acquisition. *Journal of Memory and Language*, *63*(2), 197–209. doi:10.1016/j.jml.2010.04.001

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology*, *45*(6), 1611–1617. doi:10.1037/a0016134

Vlach, H. A., & DeBrock, C. A. (2017). Remember dax? Relations between children's cross-situational word learning, memory, and language abilities. *Journal of Memory and Language*, *93*, 217–230. doi:10.1016/j.jml.2016.10.001

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375–382. doi:10.1016/j.cognition.2013.02.015

Vuong, L. C., Meyer, A. S., & Christiansen, M. H. (2016). Concurrent statistical learning of adjacent and nonadjacent dependencies. *Language Learning*, *66*(1), 8–30. doi:10.1111/lang.12137

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1), 49–63. doi:10.1016/S0163-6383(84)80022-3
Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science*, *21*(4), 221–226. doi:10.1177/0963721412449459
Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy*, *15*(4), 420–433. doi:10.1111/j.1532-7078.2009.00024.x
Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*(2), 165–180. doi:10.1111/j.1467-7687.2010.00958.x
Yu, C., & Smith, L. B. (2012a). Modeling cross-situational word – referent learning: Prior questions. *Psychological Review*, *119*(1), 21–39. doi:10.1037/a0026182
Yu, C., & Smith, L. B. (2012b). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262. doi:10.1016/j.cognition.2012.06.016
Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science*, *16*(6), 959–966. doi:10.1111/desc.12036