LANGUAGE LEARNING
A Journal of Research in Language Studies
Language Learning ISSN 0023-8333

Statistical Learning and Language: An Individual Differences Study

Jennifer B. Misyak and Morten H. Christiansen
Cornell University

Although statistical learning and language have been assumed to be intertwined, this theoretical presupposition has rarely been tested empirically. The present study investigates the relationship between statistical learning and language using a within-subject design embedded in an individual-differences framework. Participants were administered separate statistical learning tasks involving adjacent and nonadjacent dependencies, along with a language comprehension task and a battery of other measures assessing verbal working memory, short-term memory, vocabulary, reading experience, cognitive motivation, and fluid intelligence. Strong interrelationships were found among statistical learning, verbal working memory, and language comprehension. However, when the effects of all other factors were controlled for, performance on the two statistical learning tasks was the only predictor for comprehending relevant types of natural language sentences.

Keywords statistical learning; artificial grammar; language comprehension; individual differences; verbal working memory; memory span; fluid intelligence; lexical knowledge; cognitive motivation

Introduction

Statistical learning has been proposed as centrally connected to language acquisition and development. Succinctly defined as the discovery of structure by way of statistical properties of the input, such learning has been theorized to be robust and automatic and has been observed to be demonstrated across a variety of both linguistic and nonlinguistic contexts, including speech segmentation (Saffran, Aslin, & Newport, 1996), learning the orthographic and

Correspondence concerning this article should be addressed to Jennifer B. Misyak, Department of Psychology, Cornell University, 211 Uris Hall, Ithaca, NY, 14853. Internet: jbm36@cornell.edu

DOI: 10.1111/j.1467-9922.2010.00626.x

morphological regularities of written words (Pacton, Fayol, & Perruchet, 2005; Pacton, Perruchet, Fayol, & Cleeremans, 2001), learning artificial phonotactic patterns (Dell, Reed, Adams, & Meyer, 2000; Warker & Dell, 2006; Warker, Dell, Whalen, & Gereg, 2008), forming phonetic categories (Maye, Weiss, & Aslin, 2008; Maye, Werker, & Gerken, 2002), forming syntactic categories (Gerken, Wilson, & Lewis, 2005; Gómez & Lakusta, 2004), segmenting human action sequences (Baldwin, Andersson, Saffran, & Meyer, 2008), visual processing (Fiser & Aslin, 2002a, 2002b), visuomotor learning (Hunt & Aslin, 2001), tactile sequence learning (Conway & Christiansen, 2005), and nonlinguistic, auditory processing (Saffran, Johnson, Aslin, & Newport, 1999; Tillmann & McAdams, 2004). However, important issues still surround the general scope of statistical learning, especially with respect to how much of complex language structure can be captured by this type of learning.

Statistical learning research has sometimes also been studied as "artificial grammar learning" (AGL; Reber, 1967) or more broadly under the rubric of "implicit learning" (see Perruchet & Pacton, 2006). Such work has shown that infant and adult learners—upon brief and passive exposure to strings generated by an artificial grammar or continuous sequences of nonwords from an artificial lexicon—can incidentally acquire and evince knowledge for the predictive relations embedded within the stimuli (for reviews, see Gómez & Gerken, 2000; Saffran, 2003). Further, stimuli used within this paradigm may be devised so as to model structural properties specific to natural language, instantiating dependencies that may be characterized as either "adjacent" or "nonadjacent." For example, Saffran (2001) documented adults' and children's successes in incidentally learning a simplified artificial grammar that employed predictive dependencies among *adjacent* form classes (e.g., *D-E* in the string *ADE*, where each letter represents a form class defined by a set of elements). Such relationships are characteristic of natural language, in which phrasal units may be statistically signaled by dependencies between lexical members (e.g., that determiners in English predict upcoming nouns). Similarly, Gómez (2002) investigated adults' and infants' learning for an artificial grammar that generated three-element strings in which initial and final items formed a *nonadjacent* dependency pair (e.g., *a-d* of *aXd)*. Informed by the observation that certain elements in natural language belong to relatively small sets (function morphemes like *a*, *was*, *–s*, and *–ing*), whereas others belong to very large sets (open-class items such as nouns and verbs), Gómez manipulated the set size (i.e., 2, 6, 12, or 24 elements) from which she drew the middle items (*X*s), and found that participants were better able to detect the nonadjacent dependencies when the variability of the middle items was at its highest (i.e., set size 24).

Given these experimental paradigms, statistical learning appears to take place using fundamentally similar computational principles and constraints within different kinds of artificial language learning (phonological, lexical, and syntactic), across concurrent levels (e.g., the simultaneous statistical learning of lexical units and syntactic phrase structure; Saffran & Wilson, 2003), and between levels (e.g., in facilitating the mapping of subsequent lexical meanings to nonwords from a statistically segmented acoustic stream; Graf Estes, Evans, Alibali, & Saffran, 2007; Mirman, Magnuson, Graf Estes, & Dixon, 2008). Such evidence suggests that statistical learning mechanisms subserving the discovery of syntactic structure need not be distinct from those subserving the learning of nonsyntactic aspects of language such as phonology, lexicon, and semantics. However, some empirical findings have pointed to a potential distinction between forms of statistical learning that involve sequentially *adjacent* versus *nonadjacent* dependencies. Specifically, learning for these two types of dependencies have been shown to differ in their macro-level developmental trajectories and facilitative learning contexts. Within the statistical learning literature, sensitivity to nonadjacent conditional probabilities is documented later in human infancy than the earliest behavioral demonstrations of sensitivity to adjacent conditional probabilities (see Gómez & Maye, 2005, contra Saffran et al., 1996). Additionally, compared to tracking adjacent relations, most human learners generally have a harder time tracking nonadjacent dependencies (e.g., Cleeremans & McClelland, 1991; Newport & Aslin, 2004) and require more facilitative contexts to do so successfully, such as conditions that manipulate the variability of interposed items and/or exploit perceptual similarity cues (e.g., Gebhart, Newport, & Aslin, 2009; Gómez, 2002; Onnis, Christiansen, Chater, & Gómez, 2003).

This contrast between adjacent/nonadjacent statistical learning can also be seen in how researchers have typically designed studies that isolate learning for either adjacent or nonadjacent dependencies. Accordingly, the instantiation of statistical regularities among adjacent or nonadjacent stimulus tokens in these artificial grammar tasks often aims to mirror respectively the kinds of *local* or *long-distance* relations among phonemic, lexical, and phrasal constituents that individuals process in natural language. Skill in discerning both types of artificial dependencies would therefore appear relevant for many aspects of language learning, such as segmenting words and identifying phrasal boundaries (adjacent relationships) and properly inflecting morphemes and processing embeddings (nonadjacent relationships). Yet, it is unknown if these two manifestations of statistical learning are separable skills within individuals rather than denoting differing aspects of the same ability. More generally, it also remains to be fully evidenced whether and to what extent statistical learning and natural language are subserved by the same underlying mechanism(s).

The present experiment therefore employs an individual-differences framework to explore the hypothesis that statistical learning and language are integrally interrelated. The aim is to document the nature of empirical interrelationships among learner differences, informed by the observation that individual differences are substantive and ubiquitous across language. To the extent that statistical learning and language are subserved by the same underlying mechanism(s), differences in language should systematically relate to and be informative of differences in statistical learning.

Next, we briefly review findings relevant to differences in statistical learning, and then discuss the individual-difference factors of specific interest in this study.

Individual Differences in Statistical Learning

To date, findings across the statistical learning and language literatures suggest that the probabilistic knowledge resulting from statistical, implicit learning processes may substantially underpin learners' acquisition of language (e.g., for a review concerning first-language [L1] development, see Gómez, 2007; for a review that relates such effects to second-language [L2] acquisition, see N. Ellis, 2002). Whereas individual differences in language (both L1 and L2 learning/processing) have received some attention to date (for some overviews, see Bates, Dale, & Tal, 1995; Dörnyei, 2005; R. Ellis, 2004; Farmer, Misyak, & Christiansen, in press; MacDonald & Christiansen, 2002; Michael & Gollan, 2005; Vasilyeva, Waterfall, & Huttenlocher, 2008), less is known about individual differences in statistical learning within the normal population. Most evidence suggesting the presence of systematic variation in statistical learning pertains to developmental differences, atypical populations, or from studies using putative dissociations in performance between implicit and explicit learning tasks to investigate Reber's predictions (e.g., see Reber, 1993) for implicit learning as IQ-independent and age-invariant.

Thus, although seemingly present throughout development, Saffran (2001) observed consistent performance dissimilarities between children and adults in one of her artificial language studies. Additionally, Arciuli and Simpson (in press) have reported improvements in statistical learning performance as a function of increasing age in years (from 5 to 12) within typically developing children. Further, within atypical populations, performance differences on AGL or statistical learning tasks have been documented for individuals with language-related impairments: agrammatic aphasia (Christiansen, Kelly, Shillcock, & Greenfield, 2010), developmental dyslexia (Pothos & Kirk, 2004; although see counterclaims by Kelly, Griffiths, & Frith, 2002), specific language impairment (Evans, Saffran, & Robe-Torres, 2009; Hsu, Tomblin, & Christiansen, 2009), language/learning-disabled adults (Grunow, Spaulding, Gómez, & Plante, 2006; Plante, Gómez, & Gerken, 2002), and Williams syndrome children and adults (albeit not after factoring group differences in working memory or nonverbal intelligence; Don, Schellenberg, Reber, DiGirolamo, & Wang, 2003).

Finally, within the normal population, some differences in AGL have been explored in relationship to psychometric intelligence. Accordingly, Reber, Walkenfeld, and Hernstadt (1991) claimed that AGL was unrelated to intelligence, as they did not detect a significant association within their study between AGL and IQ scores from the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981), nor did McGeorge, Crawford, & Kelly (1997). However, Robinson (2005) reported a negative association between WAIS-R IQ and AGL scores in a group of experienced L2 learners. Conversely, other studies (Brooks, Kempe, & Sionov, 2006; Kempe & Brooks, 2008; Kempe, Brooks, & Kharkhurin, 2010) showed that Culture Fair Intelligence Test (CFIT; Cattell, 1971) scores mediated successful learning on miniature L2 learning tasks bearing resemblance in their design and learning demands to those invoked by traditional AGL tasks.

Therefore, although these few studies have looked at individual differences in statistical learning (sometimes with equivocal outcomes), they have not directly sought to link such differences to variations in language abilities within the normal adult population. Finding correlations between individual differences in statistical learning and language is crucial to determining whether the two may overlap in terms of their underlying mechanisms. We thus set out to explore such associations in a comprehensive study of statistical learning and language differences using a within-subject design.

Overview of Study Measures

To determine the potential role of different types of statistical learning, we used two standard artificial grammars to isolate the learning of adjacent and nonadjacent information within individuals. We then studied differences on these tasks in relation to differences in comprehending sentences whose primary manipulation entails the tracking of adjacent and/or nonadjacent natural language dependencies. As the statistical learning of adjacencies and the processing of local language dependencies both require sensitivity to adjacent sequential information, we expected that measures tapping into both of these should be more strongly interrelated than potential associations obtaining between adjacent statistical learning and the comprehension of long-distance natural language structures—and, analogously, similar expectations hold for the sensitivity to nonadjacent sequential information entailed by the statistical learning of nonadjacencies and the processing of long-distance language dependencies. Thus, the inclusion of both aspects of statistical learning allowed us to probe for any differential associations with our language measures, under the assumption that sensitivity to such dependencies is an integral component of language comprehension.

We also included in our study other potential contributing factors to variation across language and statistical learning. These measures were intended to assess differences in memory-related factors (verbal working memory and short-term memory), broad language-relevant variables (lexical knowledge and print exposure), and nonverbal abilities/aptitudes (fluid intelligence and cognitive motivation). Memory-related factors have become arguably the most studied individual-differences cognitive factors in the language processing literature and so we included them here. Verbal working memory (vWM) in particular, as conventionally gauged by reading span tasks, has been correlated with native language comprehension abilities across various experiments (for a review, see MacDonald & Christiansen, 2002). It has also begun to be extensively researched in the L2 learning literature, with studies supporting an association between L2 reading span and L2 reading skill proficiency (e.g., Harrington & Sawyer, 1992), albeit not with online processing for L2 garden-path sentences in preliminary analyses (Juffs, 2004). Research has also implicated a role for phonological short-term memory differences in L1 word learning and lexical knowledge (Baddeley, Gathercole, & Pagano, 1998) as well as in L2 acquisition (N. Ellis, 1996).

Regarding broad language-relevant variables, lexical knowledge (vocabulary) is a significant contributor to reading comprehension abilities in adolescents and adults (Baddeley, Logie, Nimmo-Smith, & Brereton, 1985; Braze, Tabor, Shankweiler, & Mencl, 2007), making it a relevant variable to account for in our study of college-aged participants. Print exposure, in turn, has been reported to be a significant predictor of lexical knowledge, even after controlling for working memory, age, and education differences (Stanovich, West, & Harrison, 1995; West, Stanovich, & Mitchell, 1993). More generally, print exposure and lexical knowledge can be considered substantial correlates for individuals' amount of reading experience, which may be logically expected to contribute to differences in reading skill. The inclusion of these two measures is therefore of potential importance in assessing the specific contribution of differences in statistical learning skills to language comprehension variance in our sample.

Finally, we incorporated two nonverbal variables in our design: fluid intelligence and cognitive motivation. Although it has been suggested that AGL is largely independent of intelligence (e.g., Reber, 1993), measures of fluid intelligence, using a nonverbal test of IQ, have been found to significantly predict individual differences on miniature L2 learning tasks (e.g., Brooks et al., 2006). We therefore included a nonverbal, fluid intelligence measure to test for an association with statistical learning performance in our tasks and to factor this variable out, as necessary, if it correlated with our statistical learning and language measures. Similarly, as motivational differences in our participants' eagerness to be engaged in demanding cognitive tasks (such as the ones employed throughout this experiment) may be a common underlying factor cutting across many of these measures, we measured cognitive motivation to control for this possibility.

Method

Participants and Materials

Thirty monolingual, native English speakers from the Cornell undergraduate population (23 women and 7 men; $M = 19.9$ years, $SD = 1.4$, range = 18–23) participated for course credit or money. To study the relationship between individual differences in statistical learning and language, we administered a test battery assessing two types of statistical learning, language comprehension, lexical knowledge, reading experience, vWM, short-term memory (STM) span, fluid intelligence (IQ), and cognitive motivation. (A summary of the tasks and measures is given in Table 1.)

Statistical Learning

Two statistical learning tasks, each implementing one of two types of artificial grammars, involving either adjacent or nonadjacent dependencies were conducted. We employed these two types of statistical learning given the possible distinction between these forms suggested by findings and approaches in the literature (see the Introduction). These types of statistical dependencies also have clear parallels within natural language structure, as sensitivity to adjacent dependencies is important for the discovery of the relationship between words within phrases and between the phrases themselves (e.g., Saffran, 2001),

Table 1 Descriptive statistics for the individual-differences tasks and measures

Task	Dependent measure	Mean (*SD*)	95% Confidence interval for mean	Observed range	Possible range
Statistical learning					
Adjacent	Percent correct (of 40 2AFC items)	62.1 (14.3)	[56.7, 67.4]	40–97.5	0–100
Nonadjacent	Percent correct (of 40 2AFC items)	69.2 (24.7)	[60.0, 78.4]	32.5–100	0–100
Language comprehension					
Animate/inanimate clauses (A/IN)	Percent correct (28 Y/N questions)	90.1 (7.2)	[87.4, 92.8]	75–100	0–100
Phonological typicality (PT)	Percent correct (20 Y/N questions)	94.4 (6.7)	[91.9, 96.9]	72–100	0–100
Subject/object relatives (S/OR)	Percent correct (40 Y/N questions)	85.6 (9.8)	[81.9, 89.3]	58–98	0–100
Other language/cognition factors					
Lexical knowledge (SILS)	Number correct (of 40) + (0.25 × number omitted)	34.4 (2.9)	[33.3, 35.5]	29–39	10–40
Reading experience (ART)	Proportion correct targets (out of 40) minus checked foils (out of 40)	0.44 (0.16)	[0.38, 0.50]	0.125–0.725	0–1
Verbal working memory (vWM)	Maximum word span with 2 of 3 trials correct (15 total trials)	4.2 (1.3)	[3.7, 4.7]	1.5–6	1–6
Short-term memory span (FDS)	Number correct trials (of 16) prior to two consecutive failures	11.0 (2.3)	[10.1, 11.9]	8–16	0–16
Fluid intelligence (CFIT)	Composite raw score (four subsections, 50 total items)	29.7 (3.6)	[28.3, 31.0]	19–36	0–50
Cognitive motivation (NFC)	Sum of scaled responses (ratings for 34 statements)	40.6 (31.6)	[28.8, 52.4]	−13– +108	−136–+136

whereas sensitivity to nonadjacent relationships between constituents is important for embeddings and long-distance dependencies (e.g., Gómez, 2002). Moreover, it has recently been suggested that different brain systems may be involved in the learning of adjacent and nonadjacent dependencies, with only the latter relevant for language (Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006).

The auditory stimuli and design structure for the statistical learning tasks were typical of those successfully used in the literature to assess statistical learning (e.g., Gómez, 2002). In particular, stimuli strings were constructed by combining individual nonword tokens recorded from a trained female, native English speaker. Assignment of particular tokens (e.g., *pel*) to particular stimulus variables (e.g., the *c* in *cXf* for the nonadjacent statistical learning task, see further below) was randomized for each participant to avoid learning biases due to specific sound properties of words. Nonwords were presented with a 250-ms interstimulus interval (ISI) within strings and a 1,000-ms ISI between strings.

For both tasks, training lasted about 25 min and was followed by a 40-item test phase. Prior to training, participants were informed that they should pay attention to the auditory sequences because they would later be tested on them, but no allusion was made to the existence of any regularities or patterns. After training, participants were informed that the sequences they just heard had been generated according to rules specifying the ordering of the nonwords. They then completed a two alternative forced choice (2AFC) test in which they were requested to discriminate grammatical strings from ungrammatical ones, with the encouragement to use "gut instinct" and impressions of familiarity/unfamiliarity to guide their judgments. Test-item pairs were presented within two blocks that counterbalanced the presentation order of grammatical and ungrammatical string versions. Half of the test pairs contained novel components that required the participant to be able to generalize the appropriate regularities to new material. These consisted of novel strings for the adjacent statistical learning task and familiar dependency pairs with novel middle elements for the nonadjacent statistical learning task. The other half of test pairs required the participant to recognize previously heard material. These involved the exact strings presented during training. Ungrammatical strings for all test-pair items differed from grammatical ones by only one element.

For the adjacent statistical learning task, the grammar was adapted with minor modification from Friederici, Steinhauer, and Pfeifer (2002) and contained adjacent dependencies occurring both within and between phrases (see Figure 1, left). Regarding phrase internal dependencies, a *D* constituent

$$S \rightarrow AP\ BP \qquad\qquad S \rightarrow \begin{Bmatrix} a\,X\,d \\ b\,X\,e \\ c\,X\,f \end{Bmatrix}$$

$$AP \rightarrow \begin{Bmatrix} D\,A \\ E\,C\,A \end{Bmatrix}$$

$$BP \rightarrow B\,(AP) \qquad\qquad X \rightarrow \{X_1, X_2, \ldots X_{24}\}$$

Figure 1 The two artificial grammars used to assess statistical learning of adjacent (left) and nonadjacent (right) dependencies.

always perfectly predicted and occurred prior to an *A* constituent, whereas an *E* constituent always directly preceded a *C* constituent that, in turn, occurred before an *A* constituent (i.e., *E C A*). Between-phrase dependencies resulted from every *B* phrase (*BP*) being consistently preceded by an *A* phrase (*AP*) and optionally followed by another *AP*. The language was instantiated through 10 distinct nonword tokens (*biv, dupp, hep, jux, lum, meep, rauk, sig, tam, zoet*) distributed over these lexical categories such that there were three *A* members, three *B* members, two *C* members, one *D* member, and one *E* member. From a set of 270 unique strings belonging to the grammar, a subset of 60 was selected as training material common to all participants and was presented in three blocks. Ungrammatical strings were produced by replacing a nonword in the string with another from a different category. For instance, if the grammatical string involved the following sequence of category constituents, *D A B D A*, a violation could entail a replacement of the second *D* with an *A*, yielding the ungrammatical string, **D A B A A* (e.g., "*jux hep lum jux biv*" *vs.* "*jux hep lum* <u>*hep*</u> *biv*"). The position of the ungrammaticality was distributed equally across categories with the exception that no violations occurred at the first or last nonword of a string (because such violations are easy to detect; Knowlton & Squire, 1996). Although strings were constructed by selecting nonwords from categories, it is important to point out that participants were exposed to all possible adjacent dependencies during familiarization. Therefore, significant discrimination by participants would reflect knowledge of adjacent structure.

Regarding nonadjacent dependencies, the ability to track relationships among remote dependencies is a fundamental linguistic ability. Previous work has shown that the statistical learning of nonadjacent dependencies is facilitated in infants and adults when there is high variability in the material that comes between the dependent elements (Gómez, 2002; Gómez & Maye, 2005; Onnis et al., 2003; Onnis, Monaghan, Christiansen, & Chater, 2004). We capitalized on this by only exposing learners to a nonadjacent dependency language incorporating high variability. Thus, for the nonadjacent statistical learning task, the grammar conformed to that of Gómez's (2002) high-variability language and consisted of three sets of dependency pairs (i.e., *a-d, b-e, c-f*), each separated by a middle *X* element (see Figure 1, right). The string-initial (*a, b, c*) and string-final (*d, e, f*) elements that comprise the nonadjacent pairings were instantiated with monosyllabic nonwords (*dak, pel, vot*; *jic, rud* and *tood*). The intervening *X*s were drawn from 24 distinct disyllabic nonwords (*balip, benez, chila, coomo, deecha, feenam, fengle, gensim, gople, hiftam, kicey, laeljeen, loga, malsig, nilbo, plizet, puser, roosa, skiger, suleb, taspu, vamey, wadim,* and *wolash*). All 72 unique sentences generated from this grammar were presented through six blocks of training. Ungrammatical strings were produced by disrupting the nonadjacency relationship with an incorrect element, thus producing strings of the form **aXe*, **bXf*, and **cXd*.

Language Comprehension

Significant differences can be found in healthy adults' ability to process sentences (see, e.g., Farmer et al., in press, for a review). We used a self-paced reading task to investigate the degree to which individual differences in language comprehension are associated with individual differences in statistical learning performance. Sentences were presented individually on a monitor using a standard word-by-word, moving window paradigm (cf. Just, Carpenter, & Woolley, 1982) and followed by "yes/no" questions probing for comprehension accuracy. Although reading times were recorded, the measures of interest for our analyses were the comprehension scores that served as offline correlates of language ability.[1] The sentence material consisted of sentences drawn from three different prior studies of various aspects of language processing (see Table 2) and chosen for this study because they entail the tracking of adjacent and/or nonadjacent dependencies in natural language. Thus, the sentence set involving clauses with animate/inanimate noun constructions (abbreviated herein as A/IN; Trueswell, Tanenhaus, & Garnsey, 1994) contained both adjacent dependencies—that is, between the animate or inanimate main clause object-noun and its modifying relative clause (e.g., *defendant/evidence* $[\ldots]_{RC}$), as well as nonadjacent dependencies holding across the relative clause, between the object-noun and the main verb (e.g., *defendant/evidence* $[\ldots]_{RC}$ *turned*). The sentence set involving noun/verb homonyms with phonologically typical or atypical noun/verb resolutions (abbreviated herein as PT; Farmer, Christiansen, & Monaghan, 2006) required tracking adjacent relations between the sentence's ambiguous homonym and the material that immediately follows

Table 2 The three language comprehension sets, with corresponding examples for each version of a given target sentence

Subject-Object Relative Clauses (S/OR)	
Subject relative:	
	The reporter that attacked the senator admitted the error.
Object relative:	
	The reporter that the senator attacked admitted the error.
Animate-Inanimate Noun Clauses (A/IN)	
Animate reduced/[unreduced]:	
	The defendant [who was] examined by the lawyer turned out to be unreliable.
Inanimate reduced/[unreduced]:	
	The evidence [that was] examined by the lawyer turned out to be unreliable.
Ambiguities involving Phonological Typicality (PT)	
Noun-like homonym with noun/verb resolution:	
	Chris and Ben are glad that the bird perches [seem easy to install]/[comfortably in the cage].
Verb-like homonym with noun/verb resolution:	
	The teacher told the principal that the student needs [were not being met]/[to be more focused].

it and locally resolves the ambiguity (e.g., *bird perches*$_{homonym}$ *seem*$_{verb}$ *vs. bird perches*$_{homonym}$ *comfortably*$_{adverb}$). The sentence set with subject-object relative clauses (abbreviated herein as S/OR; Wells, Christiansen, Race, Acheson, & MacDonald, 2009) required tracking both complex nonadjacent relationships (e.g., between the head-noun and the matrix verb across the embedded clause; *reporter [. . .]*$_{RC}$ *admitted*) and relatively simpler, more adjacent relationships (e.g., between the embedded noun and embedded verb; *senator attacked*). Four sentence lists were prepared, each incorporating 12 initial practice items, 40 sentences with subject-object relative clauses (S/OR), 28 sentences involving clauses with animate/inanimate noun constructions (A/IN), and 20 sentences involving noun/verb homonyms with phonologically typical or atypical noun/verb resolutions (PT). Sentence versions for each target sentence were counterbalanced across the four lists and presented in random order. A comprehension question was presented after each sentence. For example, after reading the last word of the sentence "*The defendant examined by the lawyer turned out to be unreliable*," the participant would press a "GO" key, which would present a new screen display with the question "*Did the lawyer question the defendant?*" After recording their response to the question by pressing either the "yes" or "no" key, participants would receive a new sentence and subsequent comprehension probe. Each participant was randomly assigned to a sentence list, and their comprehension accuracy was computed for each set of materials: S/OR, A/IN, and PT.

Lexical Knowledge

As a broad index of language skill spanning across our participants, the Shipley Institute of Living Scale (SILS) Vocabulary Subtest (Zachary, 1994), a standardized measure based on nationally representative norms, was used to assess lexical knowledge, or vocabulary. It is a paper-and-pencil measure consisting of 40 multiple-choice items in which the participant is instructed to select from among four choices the best synonym for a target word. Participants had to complete the task within 10 min.

Reading Experience

Measures of print exposure, as intended indicators of reading experience, have been found to be a significant predictor of individual differences relevant to sentence comprehension, such as vocabulary and orthographic processing (Stanovich & West, 1989; Stanovich et al., 1995). We thus used the Author Recognition Test (ART; Stanovich & West, 1989) as a traditional proxy measure of relative reading experience to assess the extent to which this may account for variance in our participants' language comprehension scores. The questionnaire required participants to check off the names of popular authors on a list. The names belonging to popular writers were chosen from a variety of print media and genres, avoiding standard school curriculum authors. The list was updated from its original form and included 40 actual authors and 40 foils. Two effort probes (the names *Edgar Allen Poe* and *Stephen King*) were also included within the list to check for attentiveness in completing the questionnaire, as these are author names that should be recognized by contemporary monolingual students attending an American college or university.

Verbal Working Memory

Differences in vWM have been associated with individual variations in sentence processing abilities (see MacDonald & Christiansen, 2002, for a review). To determine the degree to which performance on our statistical learning tasks can explain variations in sentence processing skill over and above individual differences in vWM, we used the Waters and Caplan (1996) reading span task as an assessment of our participants' vWM.[2] Participants formed yes/no semantic plausibility judgments for sets of sentences, presented one by one. At the end of a set, participants had to recall all sentence-final words in that set. The number of sentences in each set increased incrementally from two to six, with three trials at each level. Reading span was defined as the maximum level at which a participant correctly recalled all sentence-final words in two out of three trials, with no more than one failed trial at each of the preceding levels and with half of a point added if one trial had been correct at the next highest level.

Short-Term Memory Span

Whereas the above-mentioned span task is designed to measure vWM relevant for language processing, we also included an auditory Forward Digit Span (FDS) task, derived from the standardized WAIS-R subtest (Wechsler, 1981), to measure rote memory span. Among psychometric measures of individual differences in verbal short-term memory, the auditory digit span is the most widely used in the literature (Baddeley et al., 1998). A recording played a sequence of digits spoken in monotone at 1-s intervals. A standard tone after each sequence cued the participant to repeat out loud the digits they had heard in their proper order. Sequences progressed in length from two to nine digits, with two distinct sequences given for each level. Similar to WAIS-R scoring procedures, the dependent measure was the number of correctly recalled trials prior to failure on two consecutive trials.

Fluid Intelligence

General intelligence is another factor that has been suggested to affect individual differences in language and cognition (e.g., Dionne, Dale, Boivin, & Plomin, 2003). Moreover, Brooks et al. (2006) recently found that scores from the Culture Fair Intelligence Test predicted successful learning on an artificial language learning task in many ways similar to our statistical learning tasks. We therefore included this IQ test as a measure of individual differences in intelligence. We used Scale 3, Form A of the CFIT (Cattell, 1971), which is a nonverbal test of fluid intelligence or Spearman's (1927) *g*. The test contained four individually timed subsections (Series, Classification, Matrices, Typology), each with multiple-choice problems progressing in difficulty and incorporating a particular aspect of visuospatial reasoning. Raw scores on each subtest were summed together to form a composite score, which may also be converted into a standardized IQ.

Cognitive Motivation

As there may be differences across our participants in their cognitive motivation, we gauged such differences using the Need for Cognition (NFC) Questionnaire (Cacioppo & Petty, 1982) and intended to factor these out in our analyses. The NFC questionnaire provided a scaled quantification of participants' predisposition to engage in and enjoy effortful cognitive activities. Participants indicated the extent of their agreement/disagreement to 34 particular statements (e.g., "*I prefer life to be filled with puzzles that I must solve*"). We planned to examine how this measure correlates with language and statistical learning and to incorporate it as a covariate if necessary.

Procedure

Participants were individually administered the tasks during two sessions conducted on separate days (within a span of 2–9 days apart; mean interval = 5.2 days, $SD = 2.0$). For each participant, one of the two statistical learning tasks was randomly assigned for the beginning of the first session, and the other was given at the start of the second session. The remaining tasks were divided into two sets with fixed order. Set A consisted of the self-paced reading task, followed by the SILS vocabulary assessment, the NFC, and then the FDS; Set B consisted of the CFIT, the vWM span task, and then the ART. Each participant was randomly assigned one of these sets (A or B) for the first session, with the other set administered during the second session.

Results

The means, standard deviations, and range for all measures are provided in Table 1. Average performance on the two statistical learning tasks—62.1% ($SD = 14.3\%$) and 69.2% ($SD = 24.7\%$) for adjacent and nonadjacent materials,[3] respectively—was significantly above chance-level classification and indicative of learning at the group level; $t(29) = 4.63, p < .0001$ for the adjacent statistical learning task and $t(29) = 4.26, p = .0002$ for the nonadjacent statistical learning task. Each of the statistical learning tasks contained a balanced number of generalization and recognition test items (incorporating "novel" and "familiar" components respectively, as detailed under the Methods section). The average gain in accuracy for generalization items compared to recognition items was 1.2% ($SE = 2.3$) for the adjacent statistical learning task [matched pairs t test: $t(29) = 0.51, p = .61$] and was –0.8% ($SE = 2.0$) for the nonadjacent statistical learning task [matched pairs t test: $t(29) = 0.39, p = .70$]. As participants did not significantly differ in their performances on generalization and recognition tests, we collapse across these tests in subsequent analyses. Due to the experiment design, some participants received the adjacent statistical learning task during their first session ($n = 18$), whereas others received the nonadjacent statistical learning task first ($n = 12$). However, there was no main effect of statistical learning task order on participants' statistical learning scores, $F(1,28) < 1, p = .64$.

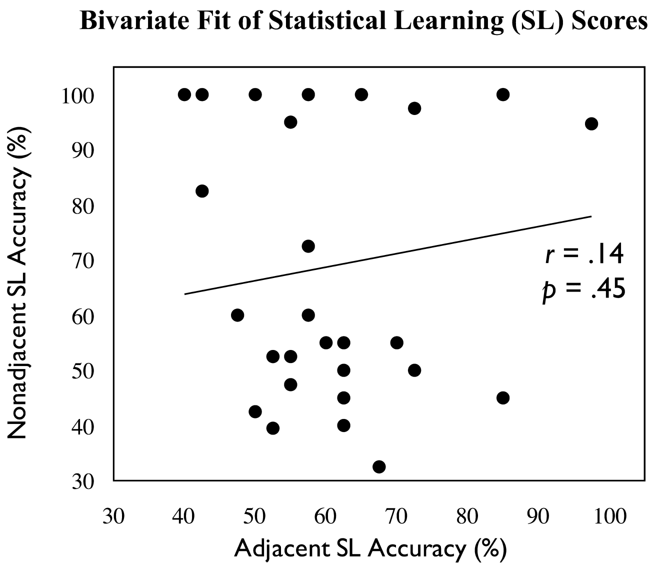

Figure 2 Participants' accuracy scores for the adjacent statistical learning (SL) task (x-axis) plotted against their accuracy scores for the nonadjacent SL task (y-axis).

The first objective in our analyses was to determine the relation between adjacent and nonadjacent dependency learning. Based on whether these correlated significantly, we intended to conduct either partial correlation analyses (in the affirmative case) or standard bivariate analyses (if no correlation was obtained). Using as our central language measures the three language scores derived from the self-paced reading task (i.e., comprehension subscores, differentiated by sentence-type),[4] we planned to explore significant correlations found between the three language measures and each of the two statistical learning measures as well as the other individual difference factors, when the effects of all measures other than a given predictor were held constant. We found no correlation between the two statistical learning tasks ($r = .14$, $p = .45$), as shown in Figure 2. We then computed the correlations between all task measures, as shown in Table 3. Regarding statistical learning, adjacent dependency learning (Adj-statistical learning) was positively associated with comprehension for the sentence set involving phonological-typicality ambiguities (PT comprehension), comprehension for the sentence set involving subject-object relative clauses (S/OR comprehension), vWM, and FDS; nonadjacent dependency learning (Nonadj-statistical learning) was positively associated with comprehension for the sentence set involving animate-inanimate noun

Table 3 Intercorrelations between task measures

	Statistical learning		Lang. comprehension			Other lang./cognition factors				
	Adjacent	Nonadj.	A/IN	PT	S/OR	SILS	ART	vWM	FDS	CFIT
NA-SL	.14									
A/IN	−.02	.41*								
PT	.49**	.12	.18							
S/OR	.39*	.42*	.11	.46*						
SILS	.05	.26	.28	.33	−.07					
ART	−.17	.16	.37*	.14	−.05	.33†				
vWM	.46*	.53**	.37*	.40*	.39*	.35†	.22			
FDS	.40*	.13	.02	.32†	.33†	.11	−.20	.36†		
CFIT	.23	.19	.20	.02	.01	.21	.07	.28	.16	
NFC	.22	.15	.33†	.32†	.03	.34†	.20	.27	.03	−.08

Note. NA-SL = Nonadj. = nonadjacent statistical learning, A/IN = animate/inanimate noun clauses, PT = ambiguities involving phonological typicality, S/OR = subject-object relative clauses, SILS = Shipley Institute of Living Scale, ART = Author Recognition Test, vWM = verbal working memory, FDS = Forward Digit Span, CFIT = Culture Fair Intelligence Test, NFC = Need for Cognition.
†$p < .09$.
*$p < .05$.
**$p < .01$ (two-tailed, $n = 30$).

clauses (A/IN comprehension), S/OR comprehension, and vWM. As can be seen in Table 3, all statistically significant correlations were of medium size, ranging between .39 and .53.

For the language-processing measures, A/IN comprehension—in addition to the positive correlation with nonadjacent statistical learning noted earlier—correlated with ART and vWM. PT comprehension, as well as correlating with adjacent statistical learning (see above), was further positively associated with S/OR comprehension and vWM. S/OR comprehension—in addition to correlating with adjacent statistical learning, nonadjacent statistical learning, and PT comprehension—correlated with vWM. Note then that there was considerable overlap in the language correlations obtained between (and among) adjacent statistical learning, nonadjacent statistical learning, and vWM. Additionally, the specific pattern of intercorrelations between statistical learning and vWM/STM indicate that adjacent statistical learning is relatively strongly associated with both vWM and STM performance (r = .46 and .40, respectively). The vWM measure is also

Table 4 Intercorrelations between language comprehension measures (left column) and statistical learning and memory-span measures (top row)

	Adj.-SL	NA-SL	vWM	FDS
A/IN	−.02	.41*	.37*	.02
PT	.49**	.12	.40*	.32†
S/OR	.39*	.42*	.39*	.33†

Note. Adj.-SL = adjacent statistical learning, NA-SL = nonadjacent statistical learning, vWM = verbal working memory, FDS = Forward Digit Span, A/IN = animate/inanimate noun clauses, PT = ambiguities involving phonological typicality, S/OR = subject-object relative clauses.
†$p < .09$.
*$p < .05$.
**$p < .01$ (two-tailed, $n = 30$).

correlated substantially with nonadjacency learning ($r = .53$), whereas STM performance only has a very weak correlation with such statistical learning ($r = .13$, $p > .09$). Thus, the kind of learning and memory skills involved in the vWM task may be more closely related to the learning of nonadjacencies than adjacencies. In contrast, the STM measure may be more closely associated with mechanisms subserving the learning of adjacent dependencies.

To ease direct comparisons between the statistical learning and memory-related measures, their intercorrelations from Table 3 are transcribed more compactly in Table 4. As evident in Table 4, vWM is well correlated with language performance in general, whereas each type of statistical learning appears to be associated more specifically with a subset of the sentence structures (as examined further in the next set of analyses).

To determine how well each measure predicted language comprehension, when controlling for all other predictors, we obtained the regression coefficient values of our individual-differences variables for each dependent language measure (Table 5). For A/IN, none of the predictors reached significance. For PT, however, Adj-statistical learning (but none of the other variables) showed a strong positive relationship to language comprehension when all other factors were held constant ($\beta = .42$, $p < .05$, one-tailed t test). For S/OR, only Nonadj-statistical learning was strongly related to language comprehension ($\beta = .38$, $p < .05$, one-tailed t test).[5] Notably, the regression coefficients for vWM were weaker (i.e., .18 or less) and much farther from reaching significance (all ps > .24). In each case then, when controlling for the effect of all other predictors, the only predictor that makes a significant and substantial contribution to the

Table 5 Regression coefficients of predictor variables for each of the dependent language comprehension measures

	Statistical learning		Other predictors					
	Adjacent	Nonadjacent	SILS	ART	vWM	FDS	CFIT	NFC
A/IN	−.20	.25	−.03	.20	.18	.01	.16	.27
PT	.42*	−.08	.24	.14	.12	.15	−.18	.07
S/OR	.28	.38*	−.20	.04	.13	.18	−.17	−.09

Note. A/IN = animate/inanimate noun clauses, PT = ambiguities involving phonological typicality, S/OR = subject-object relative clauses, SILS = Shipley Institute of Living Scale, ART = Author Recognition Test, vWM = verbal working memory, FDS = Forward Digit Span, CFIT = Culture Fair Intelligence Test, NFC = Need for Cognition.
*$p < .05$ (one-tailed, $n = 30$).

PT and S/OR language-processing measures was either of the two statistical learning measures.

Finally, we note that any comparisons between statistical learning and vWM/STM measures are clearly limited by the tasks used to assess them. If the tasks used to assess a construct are poorer than those used to assess the comparative construct(s), then the beta weights for the former predictor in regression analyses could be misleadingly attenuated in relation to the others. In principle, this could be true for vWM, although we used an established reading span task with high internal consistency and reliability (Waters & Caplan, 1996). In principle, this caveat is also applicable to our statistical learning measures; specifically in that regard, our nonadjacency task may have been prone to a ceiling effect (see Figure 2), which may have limited upper range variability in our sample and potentially reduced what would otherwise have been an even larger beta value.

Discussion

Only a few prior statistical learning studies have reported quantitative differences in performance across participants. First, with regard to adjacent statistical learning, adults performed within the range of 41–83% accuracy in the test phase of a linguistic segmentation task studied by Saffran, Newport, Aslin, Tunick, and Barrueco (1997; as reported in Evans et al., 2009). A similar range is evident upon inspection of the figures in Saffran et al. (1999), wherein the average lower bound for adult performance on nonlinguistic segmentation tasks is approximately 49% and the average upper bound is approximately

89%. The lowest and highest performances respectively across the multiple studies in Saffran et al. (1999) appear to be 33% and 97%. Although our adjacent statistical learning task involves learning an artificial grammar rather than an artificial lexicon, these observations are nonetheless consistent with our reported performance range from 40% to 97.5%.

Secondarily, with regard to our study's nonadjacent statistical learning task, which replicates the design of Gómez's (2002) high-variability condition, Gómez had noted that two thirds of learners in this condition showed perfect discrimination on a grammaticality-endorsement test measure. Analogously, we also observed perfect to near-ceiling ($\geq$95%) performances by 11 learners in this study. Although this is proportionally less than that reported by Gómez, mean performance in the original high-variability condition was also substantially higher, at 90% accuracy, than in subsequent replications for which average performance is comparable to the more modest level reported herein (see Van den Bos, Misyak, & Christiansen, 2010). Thus, although there are few documented details regarding statistical learning variation in normal adults, the variance captured by our tasks generally accords with what is known within the standard literature. This study is among the first to methodically record such information, and we encourage future researchers to include such information in their reported results.

Although it has been traditionally assumed that statistical learning processes (as commonly studied here, using AGL tasks) are largely invariant across individuals (e.g., Reber, 1993), our findings instead documented systematic variability in statistical learning performance within the normal adult population. This coincides not only with a recently emerging recognition that individual differences may exist, as even conceded in Reber and Allen (2000), but also with the development of newer paradigms intended to specifically tap into such differences (e.g., Karpicke & Pisoni, 2004; Misyak, Christiansen, & Tomblin, 2010b). As an initial investigation into these differences using a comprehensive within-subject design, our results indicated that statistical learning scores are substantially and reliably interrelated with vWM and language comprehension. Moreover, when controlling for the effects of all other predictors in the regression analyses, statistical learning ability, rather than vWM, was the only predictor of comprehension accuracy for two of the main types of sentence materials. Following MacDonald and Christiansen (2002; see also Wells et al., 2009), these results are consistent with the likely role of vWM as another index of processing skill for language comprehension and statistical learning, rather than a functionally separate capacity or mechanism. Indeed, differences in statistical learning have been recently shown to capture key online language-processing patterns previously attributed to vWM differences (Misyak et al., 2010a).

Furthermore, the specific pattern of correlations between statistical learning measures and language comprehension subscores suggests that individual differences in detecting adjacent and nonadjacent dependencies may map onto variations in corresponding skills relevant to processing similar kinds of dependencies as they occur in natural language. Thus, comprehending subject-object relative constructions in the S/OR material entails tracking long-distance relationships spanning across lexical constituents (e.g., relating the object of an embedded clause to the subject and main verb of the sentence). Analogously, the processing of items in the PT set relies upon sensitivity to adjacent information, in which the ambiguous homonym is disambiguated and locally resolved by the next word.

Because of the correlation first reported by Brooks et al. (2006) between CFIT scores and their language learning task, we had computed the correlations between CFIT and our statistical learning tasks but did not detect any significant associations. Scores, though, for nearly all of our participants were above their reported median and likely comprised a narrower range. Moreover, the lack of any associations may be consistent with subsequent findings by Gebauer and Mackintosh (2007), in which fluid intelligence correlated with AGL when participants were given detailed instructions for how to intentionally look for patterns in the training material, but not when participants were administered AGL tasks under typical instructions (as here) that promoted more incidental learning. However, our sample size combined with the probably narrower range of observed CFIT scores may have conversely limited our power to detect any potential associations.

Our experimental design included a battery of other measures that have previously received attention in studies of L1 and L2 learning, such as lexical knowledge (vocabulary), reading experience, cognitive motivation, and short-term memory span. Among these, lexical knowledge marginally correlated with print exposure (as also replicated by significant findings in Braze et al., 2007), supporting arguments for amount of reading as the best contributor to vocabulary breadth (De Temple & Snow, 2003; Krashen, 1989; Nagy & Anderson, 1984; Stanovich, 1986). The short-term memory measure (assessed via the FDS task) was found to correlate positively with adjacent, but not nonadjacent, statistical learning. Karpicke and Pisoni (2004) also had reported a correlation of equal magnitude between auditory digit span and AGL performance on an implicit sequence-learning task involving auditory or auditory-visual stimuli. Thus, the ability to recall successive elements of numerical series may covary with adjacent statistical learning skill, but this finding does not necessarily entail that a parallel relationship exists between short-term memory of phonological sequences and nonadjacency learning.

With the exception of short-term memory span and adjacency learning, then, the fact that the present study did not detect significant, strong interrelationships between most of these other variables and statistical learning performance or language comprehension does not deny their potential importance within accounts of language learning. It is furthermore possible that some of these canonical measures may be relatively weak proxies for their intended constructs (e.g., for discussion of potential limitations associated with using the ART to assess reading experience, see Acheson, Wells, & MacDonald, 2008). However, it does preliminarily support the thesis that individual differences in statistical learning skills themselves, which have been much overlooked in many explanations to date, may account for a larger proportion of language variance than the more standard measures typically used for individual differences research.

Conclusion

Overall, our findings substantiate this study's motivating rationale that variation within the normal population should provide a suitable framework for testing the empirical relatedness of language and statistical learning. As a confirmation of this approach, we found that individual differences in statistical learning exist and that sensitivity to particular kinds of statistical regularities (i.e., adjacent or nonadjacent) in the artificial grammars is predictive of processing ability for different types of sentence constructions (i.e., involving the tracking of either local or long-distance relationships). Admittedly, our study is limited by the correlational nature of its design, which cannot reveal causality, and by the relatively low number of participants, which reduces statistical power. Nonetheless, the significant results obtained here are encouraging and should be followed up by a larger scale study incorporating structural equation modeling to test these hypothesized relationships. Importantly, these findings begin to establish a heretofore missing empirical link within individuals between statistical learning and language processing (see also Conway, Bauernschmidt, Huang, & Pisoni, 2010; Misyak et al., 2010a, 2010b).

Our results may also have wider theoretical relevance to questions regarding the nature of underlying mechanism(s) for statistical learning. Although group performances for adjacent and nonadjacent grammar tasks have been documented, the research presented here is the first to assess within-subject differences across these tasks. The lack of any significant correlation detected between them, and possibly the differentiation of their predictive relations to the language measures, raises an intriguing question as to whether the two types of statistical learning may be subserved by separate mechanisms (see also Friederici et al., 2006). However, it is also possible that differences in learning strategies or task demands across the two tasks may explain the lack of association between adjacency and nonadjacency learning. Potential bimodality in the distribution of nonadjacency scores may also contribute to the lack of association. If so, this concern might be addressed with future work using newer tasks that more sensitively assess statistical learner differences (e.g., as in Misyak et al., 2010b, in which a fairly continuous and normal distribution of nonadjacency differences was documented).

The overall pattern of findings is consistent with an overlap among underlying mechanisms for both types of statistical learning and those involved for language. Counter to the claims of Friederici et al. (2006), then, we have also found that adjacency learning is substantially implicated in language for the types of sentence structures studied here. More broadly, the notion that statistical learning abilities and language may share common neurocognitive mechanisms also converges with other recent neural evidence (see, e.g., Christiansen, Conway, & Onnis, 2007; Conway & Pisoni, 2008; Friederici et al., 2002; Petersson, Forkstam, & Ingvar, 2004) and with the behavioral findings relating group differences in statistical learning to language-impaired populations (as noted within the Introduction). More research that, as here, makes within-subject comparisons across tasks is needed to understand the proper relation between different types of statistical learning and the degree to which they may be relying on the same or different neural underpinnings. Future work examining individual differences in language and statistical learning should thus aim to study in more detail the relationship between specific types of statistical structure and linguistic processing, while elucidating the nature of the underlying mechanisms upon which statistical learning and language may commonly supervene.

Revised version accepted 8 September 2010

Notes

1 Because typical AGL test measures of statistical learning, as used here, involve a substantial metacognitive component for participants' offline judgments, we considered these offline language comprehension scores to be a suitable measure for comparisons across the two tasks. Additionally, comprehension accuracy for our

participants was below ceiling, with significant variation in performance to serve as an appropriate individual-differences measure.

2 The Waters and Caplan version was used because it was reported to have greater test-retest reliability than the original Daneman-Carpenter measure (Waters & Caplan, 1996).

3 Because 5 participants received an erroneous 2AFC test-pair item on the nonadjacent statistical learning task (prior to the test pair being corrected during the course of the experiment), scores are reported as the proportion correct (with the erroneous test-pair item removed for the affected individuals). The erroneous test pair contained two ungrammatical test strings (* $a\ X_7\ e$ vs. * $b\ X_7\ d$) before being corrected to * $a\ X_7\ e$ versus $b\ X_7\ e$. None of the other test-pair items consisted of any of these specific strings (i.e., * $a\ X_7\ e$, * $b\ X_7\ d$, and $b\ X_7\ e$).

4 There were a few coding errors in the programs for presenting some of the sentence lists, resulting in the following: one fewer presented item in the sentence set corresponding to the PT manipulation for Lists 2 and 3; four fewer sentences (one S/OR item, one A/IN item, and two PT items) in List 1. List 4 was error-free. However, there was no significant effect of List on comprehension accuracy for the sentence-type sets, $F(3, 26) < 1, p = .42$.

5 As we had specifically predicted the direction of the correlation between the statistical learning and language comprehension measures to be positive, one-tailed tests were used. However, two-tailed *t* tests would still yield marginally significant correlations between adjacent statistical learning and PT comprehension ($p = .0596$) and between nonadjacent statistical learning and S/OR comprehension ($p = .0751$).

References

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods, 40*, 278–289.

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

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173.

Baddeley, A., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language, 24*, 119–131.

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition, 106*, 1382–1407.

Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *Handbook of child language* (pp. 96–151). Oxford: Basil Blackwell.

Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary: Reading skill differences in young adults. *Journal of Learning Disabilities,* 40, 226–243.

Brooks, P. J., Kempe, V., & Sionov, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied Psycholinguistics, 27*, 185–209.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116–131.

Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton-Mifflin.

Christiansen, M. H., Conway, C. M., & Onnis, L. (2007). Neural responses to structural incongruencies in language and statistical learning point to similar underlying mechanisms. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 173–178). Austin, TX: Cognitive Science Society.

Christiansen, M. H., Kelly, M. L., Shillcock, R. C., & Greenfield, K. (2010). Impaired artificial grammar learning in agrammatism. *Cognition, 116*, 382–393.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120*, 235–253.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition, 114*, 356–371.

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

Conway, C. M., & Pisoni, D. B. (2008). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Annals of the New York Academy of Sciences, 1145*, 113–131.

De Temple, J., & Snow, C. E. (2003). Learning words from books. In A. van Kleeck, S. A. Stahl, & E. B. Bauer (Eds.), *On reading books to children: Teachers and parents* (pp. 16–36). Mahway, NJ: Erlbaum.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1355–1367.

Dionne, G., Dale, P. S., Boivin, M., & Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child Development, 74*, 394–412.

Don, A. J., Schellenberg, E. G., Reber, A. S., DiGirolamo, K. M., & Wang, P. P. (2003). Implicit learning in children and adults with Williams syndrome. *Developmental Neuropsychology, 23*, 201–225.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Erlbaum.

Ellis, N. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18*, 91–126.
Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*, 143–188.
Ellis, R. (2004). Individual differences in second language learning. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 525–551). Oxford: Blackwell.
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*, 321–335.
Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103*, 12,203–12,208.
Farmer, T. A., Misyak, J. B., & Christiansen, M. H. (in press). Individual differences in sentence processing. In M. Spivey, M. Joannisse, & K. McRae (Eds.), *Cambridge handbook of psycholinguistics*. Cambridge: Cambridge University Press.
Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 458–467.
Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, 99*, 15822–15826.
Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences, 103*, 2458–2463.
Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences, 99*, 529–534.
Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 34–54.
Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review, 16*, 486–490.
Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language, 32*, 249–268.
Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431–436.
Gómez, R. L. (2007). Statistical learning in infant language development. In M. G. Gaskell (Ed.), *Oxford handbook of psycholinguistics*. Oxford: Oxford University Press.

Gómez, R. L., & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*, 178–186.

Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science, 7*, 567–580.

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

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*, 254–260.

Grunow, H., Spaulding, T. J., Gómez, R. L., & Plante, E. (2006). The effects of variation on learning word order rules by adults with and without language-based learning disabilities. *Journal of Communication Disorders, 39*, 158–170.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25–38.

Hsu, H., Tomblin, J. B., & Christiansen, M. H. (2009). *Statistical learning of nonadjacent dependencies in adolescents with and without specific language impairment*. Manuscript in preparation.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General, 130*, 658–680.

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society, 102*, 199–225.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228–238.

Karpicke, J. D., & Pisoni, D. B. (2004). Using immediate memory span to measure implicit learning. *Memory and Cognition, 32*, 956–964.

Kelly, S. W., Griffiths, S., & Frith, U. (2002). Evidence for implicit sequence learning in dyslexia. *Dyslexia, 8*, 43–52.

Kempe, V., & Brooks, P. J. (2008). Second language learning of complex inflectional systems. *Language Learning, 58*, 703–746.

Kempe, V., Brooks, P. J., & Kharkhurin, A. (2010). Cognitive predictors of generalization of Russian grammatical gender categories. *Language Learning, 60*, 127–153.

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 169–181.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal, 73*, 440–464.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*, 35–54.

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

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

McGeorge, P., Crawford, J. R., & Kelly, S. W. (1997). The relationships between psychometric intelligence and learning in an explicit and an implicit task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 239–245.

Michael, E. B., & Gollan, T. H. (2005). Being and becoming bilingual: Individual differences and consequences for language production. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*. New York: Oxford University Press.

Mirman, D., Magnuson, J. S., Graf Estes, K., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition, 108*, 271–280.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010a). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology, 1:31*. doi: 10.3389/fpsyg.2010.00031

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010b). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science, 2*, 138–153.

Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19*, 304–330.

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

Onnis, L., Christiansen, M. H., Chater, N., & Gómez, R. L. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial language learning. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Cognitive Science Society* (pp. 886–891). Boston: Cognitive Science Society.

Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Cognitive Science Society* (pp. 1047–1052). Chicago: Cognitive Science Society.

Pacton, S., Fayol, M., & Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities. *Child Development, 76*, 324–339.

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General, 130*, 401–426.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*, 233–238.

Petersson, K. M., Forkstam, C., & Ingvar, M. (2004). Artificial syntactic violations activate Broca's region. *Cognitive Science, 28*, 383–407.

Plante, E., Gómez, R., & Gerken, L. (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders, 35*, 453–462.

Pothos, E. M., & Kirk, J. (2004). Investigating learning deficits associated with dyslexia. *Dyslexia, 10*, 61–76.

Reber, A. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855–863.

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

Reber, A. S., & Allen, R. (2000). Individual differences in implicit learning: Implications for the evolution of consciousness. In R. G. Kunzendorf & B. Wallace (Eds.), *Individual differences in conscious experience* (pp. 227–247). Amsterdam: Benjamins.

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17,* 888–896.

Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition, 27*, 235–268.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language, 44*, 493–515.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*, 110–114.

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

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*, 27–52.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the comer of your ear. *Psychological Science, 8*, 101–105.

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

Spearman, C. (1927). *The abilities of man*. New York: Macmillan.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–407.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24*, 402–433.

Stanovich, K. E., West, R. F., & Harrison, M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology, 31*, 811–826.

Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1131–1142.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*, 285–318.

Van den Bos, E., Misyak, J. B., & Christiansen, M. H. (2010). *Multiple cues needed for statistical learning of probabilistic nonadjacent dependencies*. Manuscript submitted for publication.

Vasilyeva, M., Waterfall, H., & Huttenlocher, J. (2008). Emergence of syntax: Commonalities and differences across children. *Developmental Science, 11*, 84–97.

Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 387–398.

Warker, J. A., Dell, G. S., Whalen, C. A., & Gereg, S. (2008). Limits on learning phonotactic constraints from recent production experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1289–1295.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology, 49*, 51–79.

Wechsler, D. (1981). *The Wechsler Adult Intelligence Scale-Revised*. New York: Psychological Corporation.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology, 58*, 250–271.

West, R. F., Stanovich, K. E., & Mitchell, H. R. (1993). Reading in the real world and its correlates. *Reading Research Quarterly, 28*, 34–50.

Zachary, R. A. (1994). *Shipley Institute of Living Scale, revised manual*. Los Angeles: Weston Psychological Services.