Extending Statistical Learning Farther and Further: Long-Distance Dependencies, and Individual Differences in Statistical Learning and Language

Jennifer B. Misyak (jbm36@cornell.edu)
Morten H. Christiansen (mhc27@cornell.edu)
Department of Psychology, Cornell University, Ithaca, NY 14853 USA

Abstract

While statistical learning (SL) and language acquisition have been perceived as intertwined, such a view must contend with theoretical and empirical challenges. Against the backdrop of criticism leveled at early associationist efforts to account for language, a key concern for current SL approaches is whether it may suffice to enable the detection of long-distance relationships akin to those ubiquitously abounding in natural language. In Experiment 1, we extend results from previous work on the learning of nonadjacent dependencies to the learning of long-distance relations spanning three intervening elements; such learning is shown to obtain under two separate contexts. In Experiment 2, we additionally test the strength of SL and language's proposed relatedness by documenting the nature of correlations in individual differences between the two. Both experiments support the thesis that SL may overlap with mechanisms for language, while raising questions as to the singularity or duality of such underlying mechanism(s).

Keywords: Statistical Learning; Artificial Grammar Learning; Language Comprehension; Sentence Processing; Individual Differences; Working Memory; IQ

Introduction

Statistical learning (SL) 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 characterized as robust and automatic, and has been demonstrated across a variety of both linguistically relevant and general cognition contexts, including speech segmentation (Saffran, Aslin & Newport, 1996), learning the orthographic regularities of written words (Pacton, Perruchet, Fayol & Cleeremans, 2001), visual processing (Fiser & Aslin, 2002), visuomotor learning (Hunt & Aslin, 2001) and non-linguistic, auditory processing (Saffran, Johnson, Aslin & Newport, 1999). But important issues still surround the general scope of SL, especially with respect to how much of complex language structure can be captured by this type of learning.

SL research—sometimes also studied as "artificial grammar learning" (AGL) or under the rubric of "implicit learning"—has shown that infant and adult learners, upon brief and passive exposure to sequences generated by an artificial grammar, can incidentally acquire and evince knowledge for the predictive dependencies embedded within the stimuli strings (for reviews, see Gómez & Gerken, 2000; Saffran, 2003). As the instantiation of statistical regularities among stimulus tokens in such grammars commonly mirror the kinds of relations among phonemic, lexical, and phrasal constituents in actual language, a clear parallel becomes discernible between successful learning of the artificial languages and those of natural languages. Yet it remains to be fully evidenced whether and to what extent SL and language are subserved by the same underlying mechanism(s).

Furthermore, while considerable focus has been placed on the successful learning of dependencies between adjacent linguistic elements, e.g., syllables in words, comparatively less work has addressed the issue of learning nonadjacent relations (for exceptions see Gómez, 2002; Newport & Aslin, 2004; Onnis, Christiansen, Chater & Gómez, 2003). This is an area of decisive importance, as many key relationships between words and constituents are conveyed in long-distance (or remotely connected) structure. In English, for example, linguistic material may intervene between auxiliaries and inflectional morphemes (e.g., _is cooking_, _has traveled_) or between subject nouns and verbs in number agreement (e.g., _the book_s_ _on the shelf_ _are dusty_). More complex relationships to surface forms are also found in nonadjacent dependencies between antecedents and gaps, such as in _wh_-questions (e.g., _Who_ _did you see_ __?) and anaphoric reference (e.g., _John_ _went to the store where_ _he bought some apples_). Indeed, previous work incorporating statistical relations in behaviorism faulted when attempting to account for long-distance dependencies created by the presence of embedded materials. Will SL be consigned to a similar fate, found "guilty by association" or through associative shortcomings of its own? And how does SL relate to language processing more generally?

We employ a two-pronged approach to explore the hypothesis that SL and language are integrally interrelated. We first conduct an AGL experiment in which the grammar instantiates farther, more remote statistical relations than has been previously studied—thereby preliminarily testing the viability of SL to succeed in principle where earlier efforts at associationist accounts of language had been deemed to flounder. The detection of long-distance dependencies is thus the focus of our first experiment. We then press further to probe the degree to which SL and language may be empirically linked. Accordingly, our second experiment seeks to determine how individual differences in SL and natural language are interrelated.

Experiment 1: Statistical Learning of Long-Distance Dependencies

Gómez (2002) investigated 18-month-old infants' and adults' learning of nonadjacent structure by having them listen to one of two artificial languages (L1 or L2), each of

which generated three-element strings in which initial and final items formed a dependency pair (e.g., *a-d* of *aXd)*. Drawing upon 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 (nouns and verbs), and the fact that learners must often track key dependencies between functional elements, 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). Onnis et al. (2003) reported that such learning of nonadjacent relations also occurs within the visual domain and when the set size of the middle item is invariant (i.e., 1 element).

A theoretical shortcoming of these studies is that the learning of nonadjacent dependencies only occurred across a single interposed item. Here we extend those results to three middle elements. We assess for learning of the long-distance dependencies under a condition in which "overlapping" sets for the middle items of the language's generated strings (as detailed below) reflects a property of natural language in which embeddings are commonly varied and complex, admitting of interspersions in multiple places. We also include a condition without any overlap in intervening material, but which nonetheless contains the same surface-level long-distance relations as the former.

Method

Participants Thirty-nine undergraduates at Cornell University participated for course credit or monetary compensation.

Materials During training, participants listened to strings generated by an artificial language from one of two conditions. Strings in both conditions had the form *aXYZd*, *bXYZe*, and *cXYZf*, but differed in the exact composition of sets comprising the middle positions (*X, Y,* and *Z*).

In the Overlapping-Nonwords condition, $|X| = 2$, $|Y| = 3$, and $|Z| = 4$ for *aXYZd* (i.e., 2, 3, and 4 elements constituted the sets for the *X-*, *Y-*, and *Z*-positions respectively), $|X| = 3$, $|Y| = 4$, and $|Z| = 2$ for *bXYZe*, and $|X| = 4$, $|Y| = 2$, and $|Z| = 3$ for *cXYZf*. Overlap resulted from allowing four intervening elements (i.e., nonwords) to occur within two of three different sets across all dependency pairs. Using $n_1, n_2, ...n_9$ to designate the 9 distinct intervening nonwords, then the element-sets for positions *X, Y* and *Z* were as follows:

aXYZd:	*bXYZe*:	*cXYZf*:
$X = \{n_1, n_2\}$	$X = \{n_1, n_2, n_3\}$	$X = \{n_1, n_2, n_3, n_4\}$
$Y = \{n_3, n_4, n_5\}$	$Y = \{n_4, n_5, n_6, n_7\}$	$Y = \{n_5, n_6\}$
$Z = \{n_6, n_7, n_8, n_9\}$	$Z = \{n_8, n_9\}$	$Z = \{n_7, n_8, n_9\}$

Whereas in the Non-Overlap condition, $|X| = 3$, $|Y| = 3$ and $|Z| = 3$ for all three nonadjacent dependency pairings:

$$X = \{n_1, n_2, n_3\}$$
$$Y = \{n_4, n_5, n_6\}$$
$$Z = \{n_7, n_8, n_9\}$$

Strings were constructed by combining individual nonword tokens recorded from a female speaker. The initial (*a, b, c*) and final (*d, e, f*) stimulus tokens were instantiated by the nonwords *pel, dak, vot*; *rud*, *jic*, and *tood*. The middle items were drawn from the nonwords *dup, cav, jux, lum, mib, neep, tiz, rem,* and *bix*. Assignment of particular tokens (e.g., *pel*) to particular stimulus variables (e.g., the *c* in *cXYZf*) was randomized for each participant to avoid learning biases due to specific sound properties of words. Nonwords were presented with a 250 msec inter-word interval and a 750 msec inter-string interval.

Procedure Thirteen participants were recruited per condition and for a no-training control group. Since the Overlapping-Nonwords condition only had 24 unique strings for each dependency pair, 24 of 27 possible strings per dependent pair in the Non-Overlap condition were randomly selected for presentation. Trained participants listened to 4 blocks of stimuli strings, with each block composed of a random ordering of the 72 strings (24 strings x 3 pairs), for total exposure to 288 strings. Training lasted about 24 minutes.

Participants were instructed to pay attention to the stimuli because they would be tested on them later. Before testing, they were informed that the sequences they had heard were generated by a set of rules specifying the particular order of nonwords and that they would hear 12 strings, 6 of which would violate the rules. They were asked to judge whether the stimuli followed the rules by pressing a "Yes" or "No" key. Participants were then tested on a randomly ordered set of 6 grammatical strings (e.g., *aXYZd*) and 6 foils (e.g., **aXYZe*). Foils had been constructed by dissociating the tail element of a string from the string's head and replacing it with another nonword from the final-element set. Test items were identical for both conditions and for the control group.

Results and Discussion

Group means for accurate grammaticality judgments in the two conditions were statistically identical, each at 7.85 (out of 12), corresponding to 65.4% correct classification. This is significantly higher than chance-level performance, $t(12) = 2.24$, $p < .05$ for the Overlapping-Nonwords condition; $t(12) = 2.89$, $p = .01$ for the Non-Overlap condition. The control group, without any training, had a mean correct classification score of 6.31 (52.6%), which was not better than expected by chance, $t(12) = .74$, $p = .47$. Furthermore, comparisons of mean scores for each condition against that of the control's indicated significantly higher performance for both of the trained conditions: Overlapping-Nonwords versus control group, $t(24) = 1.67$, $p = .054$; Non-Overlap versus control group, $t(24) = 2.02$, $p = .028$.

While performance was modest compared to that for detecting single-item separated dependencies under high-variability contexts (cf. Gómez, 2002), significant learning was nonetheless observed. These encouraging results form a good starting point for exploring other contexts that may potentially facilitate (or hinder) detection of remote dependency structures. In support of this claim, it should be

noted that, while training duration was slightly longer than those in earlier nonadjacency studies with adults (24 minutes versus 18 minutes), participants were actually exposed to fewer total strings (288 versus 432) owing to the longer sequences (5- versus 3-element strings). Moreover, the middle-element sets' sizes were either of fairly low or of zero variability with respect to the other internal sets (i.e., versus a set size of 24). Thus, while the Overlapping condition exhibited some internal variability that may have helped place the level of learning performance on par with the Non-Overlap condition, which would be consistent with findings of Gómez (2002) and Onnis et al. (2003), languages under both conditions were learned without incorporating the facilitatory (variability) effect of large set sizes, with exposure to fewer instances of strings, and with dependent relations spanning across more items.

As a final point of interest, performance within the two conditions was seen to be highly variable across individuals. For example, 7 of the 26 trained participants (and none of the controls) demonstrated learning above 83% correct classification. Why do some individuals thus appear more adept at discerning the relevant regularities? And what implications and correspondence would such seemingly differential sensitivity to statistical structure have with respect to known population variance in natural language ability? We address these issues as part of Experiment 2.

Experiment 2: Individual Differences in Statistical Learning and Language

While individual differences in language have received some attention to date, less is known about individual differences in SL within the normal population. Although seemingly present throughout development, some minor differences across age have been documented. Saffran (2001) observed consistent performance dissimilarities between children and adults in one of her artificial language studies. Cherry and Stadler (1995) reported that SL differences, as gauged by a serial-reaction time (SRT) task, correlate with variations in educational attainment, occupational status, and verbal ability in older adults. More recently, Brooks, Kempe and Sionov (2006) showed that Culture-Fair IQ Test scores mediated successful learning on a miniature second-language learning task bearing resemblance in its design and learning demands to those invoked by a traditional AGL task. Although these few studies have looked at individual differences in SL, no previous study has directly sought to link them to variations in language abilities. Finding correlations between individual differences in SL and language is crucial to determining whether the two overlap in terms of their underlying mechanisms. We thus set out to explore this in a comprehensive study of SL and language differences using a within-subject design.

Method

Participants Thirty monolingual, native English speakers from among the Cornell undergraduate population (M=19.9,

SD=1.4) were recruited for course credit or money. None had participated in Experiment 1.

Materials To study the relationship between individual differences in SL and language, we administered a test battery assessing two types of SL, language comprehension, vocabulary, reading experience, working memory, memory span, IQ, and cognitive motivation.

Statistical Learning. Two SL tasks, each implementing one of two types of artificial grammars, involving either adjacent or nonadjacent dependencies were conducted. The auditory stimuli and design structure were typical of those successfully used in the literature to assess statistical learning (e.g., Gómez, 2002). In both tasks, training lasted about 25 minutes and was followed by a 40-item test phase. The latter used a two alternative forced choice (2AFC) format in which participants were required to discriminate grammatical strings from ungrammatical ones within sets of contrastive pairs. Ungrammatical strings differed from grammatical ones by only one element.

For the adjacent SL task, adjacent dependencies occurred both within and between phrases generated by the grammar (Figure 1, left). Regarding phrase internal dependencies, there were two types of determiners—one of which (d) always occurred prior to a noun (N), and the other of which (D) always directly preceded an adjective (A) that, in turn, occurred before a noun (D A N). Between-phrase dependencies resulted from every verb phrase (VP) being consistently preceded by a noun phrase (NP) and optionally followed by another noun phrase. The language was instantiated through 10 distinct nonwords distributed over these lexical categories such that there were 3 N, 3 V, 2 A, 1 d, and 1 D. For the nonadjacent SL task, the grammar consisted of 3 sets of dependency pairs (i.e. a-d, b-e, c-f), each separated by a middle X element (Figure 1, right). The string-initial and final elements that comprise the nonadjacent pairings were instantiated with monosyllabic nonwords. The intervening Xs were drawn from 24 distinct disyllabic nonwords. None of the nonadjacent SL nonwords were similar to those in the adjacent SL task.

S	→	NP VP	S → a X d
NP	→	d N	S → b X e
NP	→	D A N	S → c X f
VP	→	V (NP)	X = { x_1, x_2, … x_{24}}

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

Language comprehension. A self-paced reading task was used to assess language comprehension. Sentences were presented individually on a monitor using the standard moving window paradigm and followed by "yes/no" questions probing for comprehension accuracy. While 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 1). We thus computed comprehension accuracy scores for each set of materials: clauses with animate/inanimate noun constructions (A/IN; Trueswell, Tanenhaus & Garnsey, 1994), noun/verb homonyms with phonologically typical or atypical noun/verb resolutions (PT; Farmer, Christiansen & Monaghan, 2006), and subject-object relative clauses (S/OR; Wells, Christiansen, MacDonald & Race, 2007).

Table 1: Language comprehension sentence examples.

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)

Reduced: *The defendant/evidence examined by the lawyer turned out to be unreliable.*

Unreduced: *The [defendant who]/[evidence that] was examined by the lawyer turned out to be unreliable.*

Ambiguities involving Phonological Typicality (PT)

Noun-like homonym with N/V resolution: *Chris and Ben are glad that the bird perches [seem easy to install]/[comfortably in the cage].*

Verb-like homonym with N/V resolution: *The teacher told the principal that the student needs [were not being met]/[to be more focused].*

Vocabulary. The Shipley Institute of Living Scale (SILS) Vocabulary Subtest (Zachary, 1994) was used to assess 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.

Reading Experience. The Author Recognition Test (ART) (Stanovich & West, 1989) was used as a proxy measure of relative reading experience. The questionnaire required participants to check off the names of popular writers they recognize on a list. The list included 40 actual authors, 40 foils, and 2 "effort probes."

Working Memory. The Waters and Caplan (1996) reading span task gauged verbal working memory (vWM). Participants were asked to recall all sentence-final words of a given sentence set, while forming semantic judgments for each individual sentence as it was visually presented. The number of sentences in each set increased incrementally from 2 to 6, with three trials at each level.

Memory Span. Rote memory capacity was indexed through recall accuracy on the Forward Digit Span (FDS) task, derived from the WAIS-R subtest (Wechsler, 1981). A recording played a sequence of digits spoken in monotone at

---

[1] Given the offline nature of SL grammaticality tests, these offline comprehension measures are more suitable for comparisons than simple RTs (reading times) as they better equate task demands across the experimental manipulations.

1-sec 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 2 to 9 digits, with two distinct sequences given for each level.

IQ. We used Scale 3, Form A of Cattell's Culture Fair Intelligence Test (CFIT) (1971), which is a nonverbal test of fluid intelligence or Spearman's *"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 are summed together to form a composite score, which may also be converted into a standardized IQ.

Cognitive Motivation. The Need for Cognition (NFC) Questionnaire (Cacioppo, Petty & Kao, 1984) provided a scaled quantification of participants' disposition 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."*).

Procedure Participants were individually administered the tasks during two sessions on separate days. For each participant, one of the two SL tasks was randomly assigned for the beginning of the first session, and the other was given at the start of the second session. In addition to these tasks for assessing statistical learning, participants completed the measures of language and cognitive factors noted above: self-paced reading task, SILS vocabulary assessment, ART, reading span task, FDS, CFIT, and NFC.

Results and Discussion

The mean performance on the two SL tasks—62.1% (*SD*=14.3%) and 69.2% (*SD*=24.7%) for adjacent and nonadjacent respectively—was significantly above chance-level classification and indicative of learning at the group-level; $t(29) = 4.63$, $p < .0001$ for the adjacent SL task; $t(29) = 4.26$, $p = .0002$ for the nonadjacent SL task. The means for the other measures were as follows: A/IN (*M*=90.1%, *SD*=7.2%), PT (*M*=94.4%, *SD*=6.7%), S/OR (*M*=85.6%, *SD*=9.8%), SILS (*M*=34.4, *SD*=2.9), ART (*M*=0.44, *SD*=0.16), vWM (*M*=4.2, *SD*=1.3), FDS (*M*=11.0, *SD*=2.3), CFIT (*M*=29.7, *SD*=3.6), and NFC (*M*=40.6, *SD*=31.6).

The first objective in our analyses was to determine the relation between adjacency 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), we planned to explore significant correlations found between the three language measures, the two SL measures, and the other individual difference factors, using stepwise regressions with Bonferoni corrections for multiple comparisons.

We found no correlation between the two SL tasks ($r = .14$, $p = .45$). We then computed the correlations between all

Table 2: Intercorrelations between task measures in Experiment 2.

	Adj-SL	Nonadj-SL	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

†$p < .09$. *$p < .05$. **$p < .01$. (two-tailed, $n = 30$).

task measures as shown in Table 2. Regarding SL, adjacent dependency learning ("Adj-SL") was positively associated with PT-comprehension ("PT-comp"), S/OR-comp, vWM, and FDS; nonadjacent dependency learning ("Nonadj-SL") was associated with A/IN-comp, S/OR-comp, and vWM.

For the language-processing measures, A/IN-comp—in addition to the positive correlation with Nonadj-SL noted above—correlated with ART and vWM. PT-comp, as well as correlating with Adj-SL (above), was further positively associated with S/OR-comp and vWM. And S/OR-comp—besides correlating with Adj-SL, Nonadj-SL, and PT-comp—correlated with vWM. Note then that there was considerable overlap in the language correlations obtained between (and among) Nonadj-SL, Adj-SL, and vWM.

To determine which of our measures were the best predictors of language comprehension and whether other measures would explain part of the variance in those scores after entry of each corresponding score's strongest predictor, we carried out three stepwise regression analyses. The variables from the bivariate analyses that were significant at the .05 level were entered as predictors and the language comprehension scores entered as the dependent variables (P value for entry = .05, P value for remaining = .10). The stepwise regression for A/IN-comp revealed only a single variable in the model: Nonadj-SL, $t(29) = 2.39$, $p = .024$, $R^2 = .17$ (i.e., ART and vWM did not enter). After regression for PT-comp, the only variable left in the model was Adj-SL, $t(29) = 2.96$, $p = .006$, $R^2 = .24$. And for S/OR-comp, Nonadj-SL alone predicted the scores after regression, $t(29) = 2.47$, $p = .020$, $R^2 = .18$. In each case then, the best (and sole) predictor of the language-processing measure was either of the two SL measures.

Because of the correlation reported by Brooks et al. (2006) between CFIT (IQ) scores and their language-learning task, we computed the correlations between CFIT and our SL tasks, but did not detect any significant associations; however, scores for nearly all our participants were above their reported median and likely comprised a narrower range. We also note that Vocabulary, traditionally construed as a relative proxy for language experience, did not correlate with the SL tasks, but did correlate with marginal significance to PT-comp ($p = .073$), ART ($p = .076$), NFC ($p = .069$), and vWM ($p = .055$) factors.

Our findings confirmed systematic variability in SL performance across the normal adult population, and indicated that SL scores were also strongly interrelated with vWM and language comprehension. Moreover, SL ability, rather than vWM, was the single best predictor of comprehension accuracy for each of the types of sentence material in the regression models. Following MacDonald and Christiansen (2002), these results are consistent with the likely role of vWM as merely another index of processing skill for language comprehension and SL, rather than a functionally separate mechanism.

Furthermore, the specific pattern of correlations between SL 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, statistics underlying successful processing of A/IN material also invoke long-distance elements given the ambiguous nature and relative clause construction common to most sentences of that set. And while the nature of individual differences in the processing of phonologically typical lexical items has yet to be fully known, it seems plausible that individual sensitivity to such cues relies upon detecting sequential phonological regularities that are in essence adjacently co-occurring and thus hinge upon attunement to local (i.e., adjacent) relations.

General Discussion

As language often involves interspersing several words or linguistic constituents between long-distance dependencies, it is critical to determine the extent to which this can be accomplished via SL. The results in Experiment 1 build upon the formative findings of Gómez (2002), who studied the learning of nonadjacent structure in three-element dependency strings, and extend them to five-element sequences—showing that the detection of farther, surface-level long-distance statistical relationships than previously reported is, in fact, possible by human learners. Additionally, sensitivity to such statistical structure was

demonstrated across two different circumstances (i.e., variations in the permissibility of overlapped words and in the relative set sizes of the intervening material). These findings are a step forward in scaling up the complexity of artificial grammars to match with that of language, and aid in preliminarily countering concerns for the feasibility of current SL approaches to account for the learning of long-distance relationships common to language—a point of contention besetting earlier behaviorist endeavors.

Experiment 2 shows that the variation in learning performance observed in Experiment 1 also pertains to SL tasks instantiating more standard grammars and that such variation within the normal population may provide a suitable framework for further testing the empirical relatedness of language and statistical learning. As a confirmation of this approach, it appears that sensitivity to particular kinds of statistical regularities (i.e., adjacent or nonadjacent) in the artificial grammars was predictive of processing ability for different types of sentence constructions (i.e., involving the tracking of either local or long-distance relationships).

Our results may also be relevant to questions regarding the nature of underlying mechanism(s) for SL. 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 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 SL may be subserved by separate mechanisms. More research that, as here, makes within-subject comparisons across tasks is needed to understand the proper relation between different types of SL and the degree to which they may be relying on the same or different neural underpinnings.

Further work in the learning of long-distant dependencies, in tandem with examining individual differences in language and statistical learning, should thus aid in mapping more concretely the relation between statistical sensitivities and linguistic processing, while elucidating the nature of the underlying mechanism(s) upon which statistical learning and language may commonly supervene.

Acknowledgments

References

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., Petty, R., & Kao, C. (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.

Cherry, K.E., & Stadler, M.A. (1995). Implicit learning of a nonverbal sequence in younger and older adults. *Psychology and Aging, 10,* 379-394.

Farmer, T.A., Christiansen, M.H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103,* 12203-12208.

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

Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science, 13,* 431-436.

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

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(4),* 658-680.

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.

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. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial language learning. *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 886-891). Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Saffran, J.R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12(4),* 110-114.

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

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.

Stanovich, K.E., & West, R.F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24(4),* 402-433.

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.

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.

Wells, J., Christiansen, M.H., MacDonald, M.C., & Race, D. (2007). *Experience and sentence comprehension: Statistical learning, working memory, and individual differences.* Submitted manuscript.

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

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