

Original citation:

Misyak, Jennifer. (2010) On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, Volume 1 (Number 31).
<http://dx.doi.org/10.3389/fpsyg.2010.00031>

Permanent WRAP url:

<http://wrap.warwick.ac.uk/68105>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution 4.0 International license (CC BY 4.0) and may be reused according to the conditions of the license. For more details see: <http://creativecommons.org/licenses/by/4.0/>

A note on versions:

The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: publications@warwick.ac.uk

warwick**publications**wrap

highlight your research

<http://wrap.warwick.ac.uk>



On-line individual differences in statistical learning predict language processing

Jennifer B. Misyak¹, Morten H. Christiansen^{1*} and J. Bruce Tomblin²

¹ Department of Psychology, Cornell University, Ithaca, NY, USA

² Department of Communication Sciences and Disorders, University of Iowa, Iowa City, IA, USA

Edited by:

Gabriella Vigliocco, University College London, UK

Reviewed by:

Michael Spivey, University of California at Merced, USA

Gary Dell, University of Illinois at Urbana-Champaign, USA

*Correspondence:

Morten H. Christiansen, Department of Psychology, Cornell University, 228 Uris Hall, Ithaca, NY, 14853, USA.
e-mail: christiansen@cornell.edu

Considerable individual differences in language ability exist among normally developing children and adults. Whereas past research have attributed such differences to variations in verbal working memory or experience with language, we test the hypothesis that individual differences in statistical learning may be associated with differential language performance. We employ a novel paradigm for studying statistical learning on-line, combining a serial-reaction time task with artificial grammar learning. This task offers insights into both the timecourse of and individual differences in statistical learning. Experiment 1 charts the micro-level trajectory for statistical learning of nonadjacent dependencies and provides an on-line index of individual differences therein. In Experiment 2, these differences are then shown to predict variations in participants' on-line processing of long-distance dependencies involving center-embedded relative clauses. The findings suggest that individual differences in the ability to learn from experience through statistical learning may contribute to variations in linguistic performance.

Keywords: statistical learning, language processing, individual differences, serial-reaction time, artificial grammar, relative clauses

INTRODUCTION

Individual differences are ubiquitous and substantial across language development and use, prompting much debate regarding the underlying sources for this variation (see Bates et al., 1995; MacDonald and Christiansen, 2002, for reviews). As in other areas of psychology, the relative importance of biological versus experiential factors has figured prominently in these discussions. Thus, a "capacity-based" viewpoint has attributed inter-individual variability to constraints on cognitive resources or capacities, such as limitations arising from an individual's working memory (e.g., Just and Carpenter, 1992; cf. Waters and Caplan, 1996). An alternative, "experience-based" account instead has highlighted the role of experiential factors in shaping linguistic skills (e.g., MacDonald and Christiansen, 2002; Wells et al., 2009). Here, we pursue the hypothesis that individual differences in the ability to learn from experience by way of statistical learning contribute to variations in language performance.

Studies of statistical learning have shown that humans are sensitive to various statistical aspects of their environments, incidentally learning not only about the simple frequency of events (Hasher and Zacks, 1984) and their co-occurrences (Kirkham et al., 2002) but also of conditional regularities obtaining across a variety of perceptual, social, and cognitive contexts – from the processing of visual scenes (Fiser and Aslin, 2002a) to segmenting human action sequences (Baldwin et al., 2008) to identifying the word boundaries in running speech (Saffran et al., 1996a) and discovering predictive syntactic relationships (Saffran, 2002). Crucially, statistical learning has been proposed to be a key mechanism for acquiring knowledge of probabilistic dependencies intrinsic to linguistic structure (see Gómez and Gerken, 2000; Saffran, 2003,

for reviews). Thus, variations in statistical learning may be a mediating factor that affects an individual's ability to learn from linguistic experience.

Speaking against this hypothesis is the general assumption that statistical learning is invariant across individuals, development, and psychological disorders (e.g., Reber, 1993; Cleeremans et al., 1998), with many studies showing remarkable similarities in infant, child, and adult learning (e.g., Cherry and Stadler, 1995; Saffran et al., 1996a,b; Thomas and Nelson, 2001; Fiser and Aslin, 2002a,b; Gómez, 2002; Saffran, 2002). However, considerable individual differences in statistical learning do exist, and tend to be associated with differences in language ability. For example, children with specific language impairment (SLI) have problems with statistical learning in a visual sequence learning task (e.g., Tomblin et al., 2007). Additionally, substantial differences in statistical learning performance have been found even within the normal adult population (Misyak and Christiansen, in press), suggesting that systematic differences in statistical learning itself may be a largely overlooked contributor to variations in language performance.

To investigate this possibility, we employed a novel paradigm (see Misyak et al., 2010, for further detail), which combines advantages of both conventional artificial grammar learning (AGL; Reber, 1967) and serial reaction time (SRT; Nissen and Bullemer, 1987) approaches, to examine statistical learning on-line, and we then applied our task towards studying the acquisition and processing of nonadjacent dependencies. As language abounds in long-distance dependencies that learners must track on-line (e.g., subject-verb agreement, clausal embeddings, and relationships between auxiliaries and inflected morphemes), an increasing body of statistical learning work has been directed towards examining nonadjacency

learning (e.g., Gómez, 2002; Onnis et al., 2003; Newport and Aslin, 2004; Lany and Gómez, 2008; Pacton and Perruchet, 2008; Gebhart et al., 2009). Importantly, the ability to track long-distance dependencies in natural language has also figured prominently in the debate on individual differences in adult sentence processing. Experiment 1 therefore implements Gómez's (2002) artificial nonadjacency language within the new hybrid "AGL-SRT" task to uncover the timecourse of learning and establish a sensitive index of inter-individual differences. Experiment 2 then links learning performance on this new task to variations in the same individuals' on-line processing of long-distance dependencies in natural language sentences with embedded relative clauses.

EXPERIMENT 1: ON-LINE STATISTICAL LEARNING OF NONADJACENT DEPENDENCIES

Nonadjacent statistical learning has been well-documented under "high variability" contexts, whereby a relatively larger set size from which a middle string-element is drawn facilitates learning of the nonadjacent relationship between the two flanking elements (Gómez, 2002). In other words, when exposed to artificial, auditory strings of the form aXd and bXe , infants and adults display sensitivity to the nonadjacencies (i.e., the a_d and b_e relations) when X is drawn from a large set of exemplars (e.g., when $|X| = 18$ or 24). Performance is impeded, however, when variability among X -elements is lowered for smaller set sizes (e.g., $|X| = 12$ or 2). Although subsequent studies have obtained similar results with visual stimuli (Onnis et al., 2003) and highlighted the importance of prior experience (Lany and Gómez, 2008), little is known about the timecourse of high-variability nonadjacency learning as it actually unfolds (though see Misyak et al., 2010). Here, we address this gap by using the novel AGL-SRT paradigm to reveal group patterns and individual differences in corresponding learning trajectories.

METHOD

Participants

Fifty native English speakers from the Cornell undergraduate population (age: $M = 19.8$, $SD = 1.5$) participated for course credit or \$10. All participants provided their informed consent and the Cornell Institutional Review Board has approved both experiments reported here.

Materials

During training, participants were exposed cross-modally to strings from Gómez's (2002) artificial high-variability, nonadjacency language. Strings had the form aXd , bXe , and cXf , with initial and final items forming a dependency pair. Beginning and ending stimulus tokens (a , b , c ; d , e , f) were instantiated by the nonwords *pel*, *dak*, *vot*, *rud*, *jic*, and *tood*; middle X -tokens were instantiated by 24 disyllabic nonwords: *wadim*, *kicey*, *puser*, *fengle*, *coomo*, *loga*, *gople*, *taspu*, *hifam*, *deecha*, *vamey*, *skiger*, *benez*, *gensim*, *feenam*, *laeljeen*, *chila*, *roosa*, *plizet*, *balip*, *malsig*, *suleb*, *nilbo*, and *wiffle*. Assignment of particular tokens (e.g., *pel*) to particular stimulus variables (e.g., the c in cXf) was randomized for each participant to avoid learning biases due to specific sound properties of words. Mono- and bi-syllabic nonwords were recorded with equal lexical stress from a female native English speaker and length-edited to 500 and 600 ms respectively. Ungrammatical items were produced by disrupting

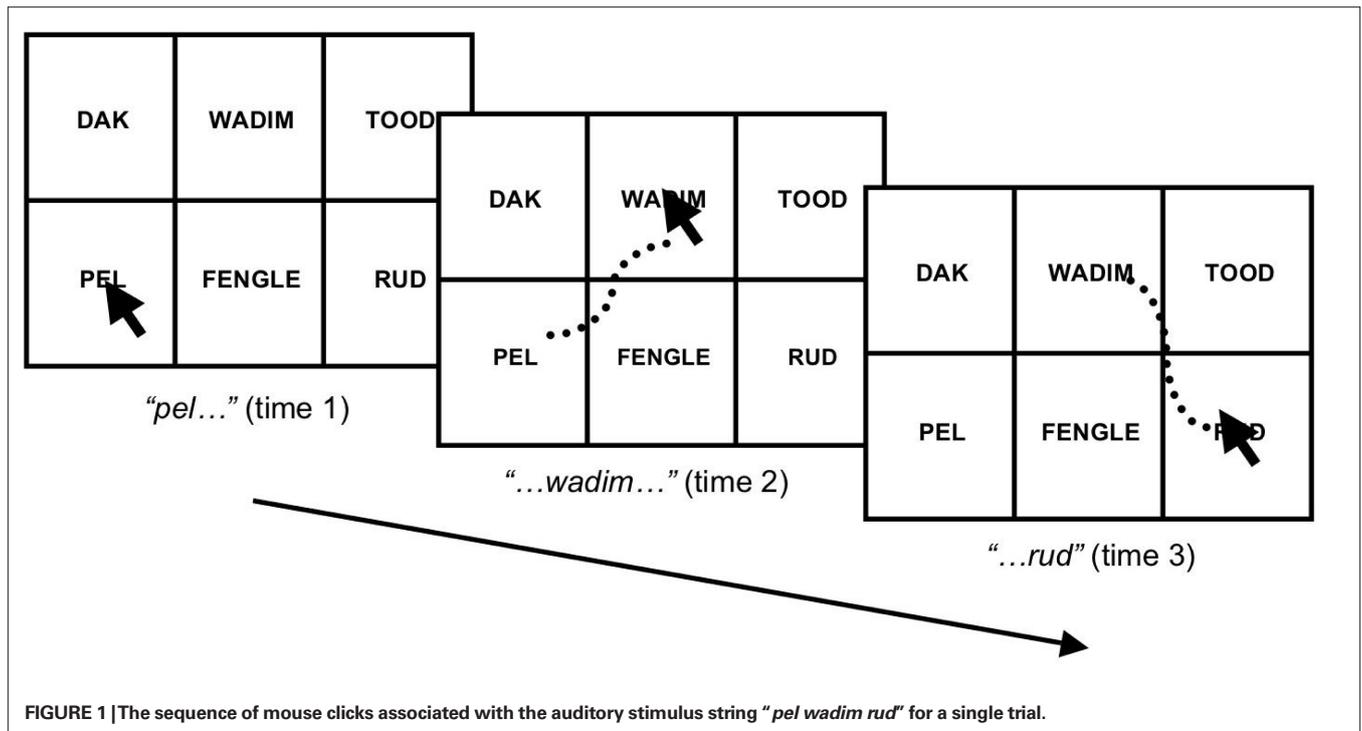
the nonadjacent relationship with an incorrect final element to produce strings of the form: $*aXe$, $*aXf$, $*bXd$, $*bXf$, $*cXd$, and $*cXe$. Written forms of nonwords (in Arial font, all caps) were presented using standard spelling.

Procedure

A computer screen was partitioned into a grid consisting of six equal-sized rectangles: the leftmost column contains the beginning items (a , b , c), the center column the middle items ($X_1 \dots X_{24}$), and the rightmost column the ending items (d , e , f). Each trial began by displaying the grid with a written nonword centered in each rectangle, with each column containing a nonword from a correct and an incorrect stimulus string (foils). Positions of the target and foil were randomized and counterbalanced such that each occurred equally often in the upper and lower rectangles. Foils were only drawn from the set of items that can legally occur in a given column (beginning, middle, end). For example, for the string *pel wadim rud* the leftmost column might contain *PEL* and the foil *DAK*, the center column *WADIM* and the foil *FENGLE*, and the rightmost column *RUD* and the foil *TOOD*, as shown in **Figure 1** across three time steps.

After 250 ms of familiarization to the six visually presented nonwords, the auditory stimuli were played over headphones. Participants were instructed to use a computer mouse to click upon the rectangle with the correct (target) nonword as soon as they heard it, with an emphasis on both speed and accuracy. Thus, when listening to *pel wadim rud* the participant should first click *PEL* upon hearing *pel* (**Figure 1**, left), then *WADIM* when hearing *wadim* (**Figure 1**, center), and finally *RUD* after hearing *rud* (**Figure 1**, right). After the rightmost target has been clicked, the screen clears, and a new set of nonwords appears after 750 ms. An advantage of this design is that every nonword occurs equally often (within a column) as target and as foil. This means that for the first two responses in each trial (leftmost and center columns), participants cannot anticipate beforehand which is the target and which is the foil. Following the rationale of standard SRT experiments, however, if participants learn the nonadjacent dependencies inherent in the stimulus strings, then they should become increasingly faster at responding to the final target. The dependent measure is therefore the reaction time (RT) for the predictive, final element on each trial, subtracted from the RT for the nonpredictive, initial element, which serves as a baseline and control for practice effects.

Each training block involved the random presentation of 72 unique strings (24 strings \times 3 dependency-pairs). After exposure to these 432 strings (across the first six training blocks), participants were presented with 24 ungrammatical strings, with endings that violated the dependency relations (as noted above). The inclusion of a continuous block of ungrammatical items is roughly analogous to SRT designs interposing a block of random violations between blocks of structured sequences (e.g., Thomas and Nelson, 2001). In contrast to randomly interspersing violations throughout all blocks, this design is suitable here given (a) the relatively small number of overall trials, (b) our aim to obtain a clear temporal trajectory of nonadjacency learning, and (c) our extraction of an on-line learning metric unaffected by the later introduction of ungrammatical items (see "Results and Discussion"). This short ungrammatical block (i.e., with two-thirds fewer items than a

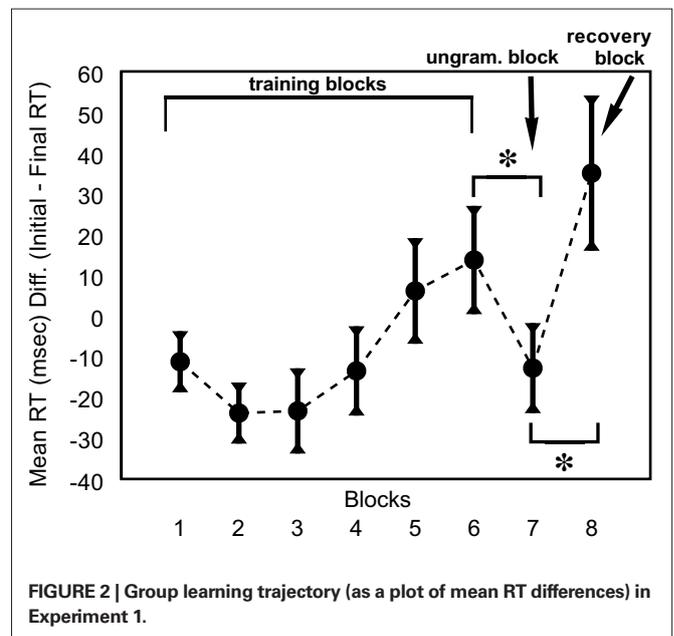


training block) was next followed by a final (“recovery”) block with 72 grammatical strings. Block transitions were seamless and unannounced to participants.

After the presentation of all eight blocks, participants completed a standard grammaticality test often used to assess statistical learning within AGL designs (e.g., Gómez, 2002). Participants were informed that the sequences they heard had been generated according to rules specifying the ordering of nonwords, and that they would hear 12 strings, six of which would violate the rules. They were instructed to endorse or reject each string according to whether they judged it to follow the rules. Participants were presented with a randomly ordered set of six grammatical strings (e.g., *aXd*) and six foils (e.g., **aXe*). Foils were produced in the same manner as ungrammatical items for the AGL-SRT task, with the exception that none of the exact foil strings at test (e.g., **aX₄e*) had occurred in the ungrammatical block.

RESULTS AND DISCUSSION

Since participants were instructed to respond both as accurately and as quickly as possible, a small number of errors was observed. Analyses were performed on only accurate string trials (with no more than one selection response for each of the three targets). This criterion is quite conservative, as standard SRT designs typically consider accuracy with respect to single-selection responses defining one “trial,” rather than for all three string-elements composing a string-trial in our design. Accordingly, accurate string-trials for our data analyses comprised grand averages of 91.4% (SD = 5.7) of training block trials, 89.8% (SD = 9.5) of ungrammatical trials, and 90.0% (SD = 8.1) of recovery trials. [In comparison, selection accuracy for the single final-element across trial-types was 96.3% (2.8), 95.0% (5.4), and 95.7% (4.6).]



Mean RT difference scores (i.e., initial-element RT minus final-element RT) were computed for each block and submitted to a one-way repeated-measures analysis of variance (ANOVA) with block as the within-subjects factor. As Mauchly’s test indicated a violation of the sphericity assumption ($\chi^2(27) = 141.96, p < 0.001$), degrees of freedom were corrected using Greenhouse–Geisser estimates ($\epsilon = 0.51$). Results indicated that mean RT difference was affected by block, $F(3.58, 175.28) = 5.59, p = 0.001$. **Figure 2** plots mean RT difference scores averaged within blocks,

with improvements from baseline (block 1 performance) reflecting nonadjacency learning. RT differences gradually increased throughout, albeit with an expected decline in the ungrammatical seventh block. Cleeremans and McClelland (1991) have previously found that sensitivity to long-distance contingencies emerges more gradually than for adjacent dependencies; our temporal trajectory in **Figure 2** also indicates that sensitivity to nonadjacent dependencies requires considerable exposure (five blocks on average) before it reliably affects responses.

Planned contrasts confirmed that mean RT differences in the ungrammatical block significantly decreased compared to both the preceding training block, $t(49) = 2.27, p = 0.03$, and the subsequent recovery block, $t(49) = 3.06, p = 0.004$. Following interpretations in the implicit learning literature for comparing RTs to structured (patterned) versus unstructured (random) material (e.g., Thomas and Nelson, 2001), this decrement in performance (Block 6 minus Block 7: $M = -26.5$ ms, $SE = 11.7$) provides evidence for participants' sensitivity to violations of the sequential structure, with a return to improved performance demonstrated upon the reinstatement of grammatical sequences in the recovery block (Block 8 minus Block 7: $M = 47.8$ ms, $SE = 15.6$). Although recovery performance was nominally higher than training performance in Block 6 (with a mean difference of 21 ms), potentially reflecting individuals' continued learning throughout the duration of the last block, participants' RTs in the recovery block exhibited greater variation and differed only marginally from Block 6 performance ($t(49) = 1.69, p = 0.10$).

Performance on the standard grammaticality judgment test averaged 58.2% ($SD = 17.1$) with substantial inter-individual variation. This group-level performance was above chance, ($t(49) = 3.37, p = .0015$), providing an off-line confirmation of nonadjacency learning using a standard measure of statistical learning. Additionally, we calculated an on-line learning score for each participant by subtracting their RT performance in the first training block (Block 1) from that in the final training block (Block 6), with positive values indicating pattern-specific learning across the six blocks of training (see Cherry and Stadler, 1995, for a similar approach). Averaged across participants, the on-line learning scores were significantly above zero, with large individual differences ($M = 24.9$ ms, $SD = 87.6$), $t(49) = 2.01, p = 0.05$. This measure of on-line learning also correlated positively with the speed-up found for the recovery block (Block 8 minus Block 7: $r = 0.39, p = 0.005$). Thus, the on-line learning score provides a reliable index of differences in participants' sensitivity to the nonadjacencies in the AGL-SRT task.

In sequence learning paradigms, participants' knowledge may be more robustly evidenced – and sometimes exclusively expressed – through indirect measures (i.e., reaction-times), rather than through more direct assessments that may rely on metaknowledge (e.g., see Jiménez et al., 1996). Our on- and off-line measures (RT and grammaticality judgments, respectively) did not correlate with one another ($r = 0.15, p = 0.30$), which resonates with SRT findings that more implicit/indirect versus more explicit/direct performance measures may be functionally dissociable (Willingham et al., 1989; Cohen et al., 1990; see also Destrebecqz and Cleeremans, 2001). We therefore use our indirect RT measure (the Block 6 minus Block 1 performance difference)

to index individual differences in statistical learning, with the added advantage that this allows for on-line comparisons between learning/processing measures across Experiments 1 and 2.

To further investigate the learning trajectories of good and poor statistical learners, we grouped participants based on their on-line learning scores, with values above zero distinguishing “good” ($n = 28, M = 78.7$ ms, $SE = 14.9$) from “poor” ($n = 22, M = -43.5$ ms, $SE = 7.1$) learners. Both groups' temporal processing patterns for nonadjacencies are plotted in **Figure 3**. Inspection of these trajectories reveals distinct group differences concerning the overall shape of the statistical learning trajectory and the response to ungrammatical items. Notably, the critical block contrasts for demonstrating nonadjacency learning at the group-level were also significant within the subgroup of good learners (Block 7 minus Block 6: $M = 42.3$ ms, $SE = 18.4, t(27) = 2.29, p = 0.03$; Block 8 minus Block 7: $M = 60.3$ ms, $SE = 24.8, t(27) = 2.24, p = 0.02$). However, the poor statistical learners showed little evidence of learning (Block 7 minus Block 6: $M = -6.40$ ms, $SE = 11.6, t(21) = 0.55, p = 0.59$; Block 8 minus Block 7: $M = 32.0$ ms, $SE = 16.5, t(21) = 1.93, p = 0.07$). The next experiment further looks at the consequences of these differences in AGL-SRT learning for individuals' on-line language processing.

EXPERIMENT 2: ON-LINE INDIVIDUAL DIFFERENCES IN LANGUAGE PROCESSING AND STATISTICAL LEARNING

Individual differences in tracking long-distance dependencies in natural language have been extensively studied in relation to the contrastive processing of subject and object relative clauses. Object relative (OR) sentences (illustrated in 2) involve a head-noun that is the object of an embedded clause, and are generally more difficult to process and comprehend than subject relatives (SRs; such as 1), in which the head-noun is the subject of the modifying clause (though see Real and Christiansen, 2007).

- (1) *The reporter that attacked the senator admitted the error.*
- (2) *The reporter that the senator attacked admitted the error.*

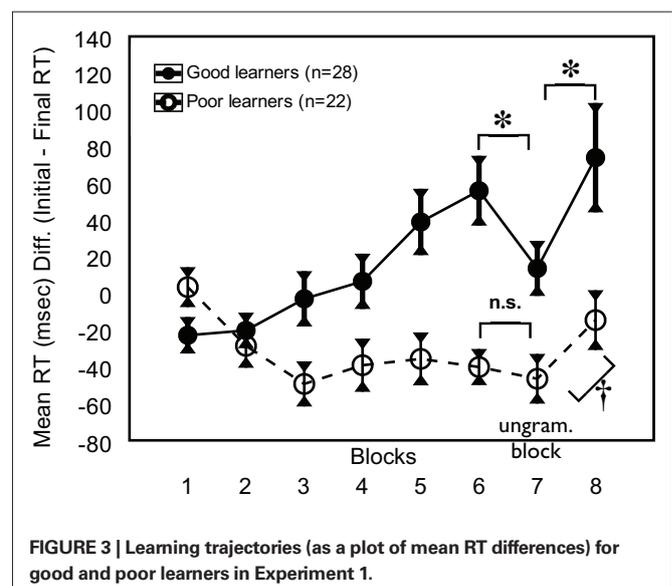


FIGURE 3 | Learning trajectories (as a plot of mean RT differences) for good and poor learners in Experiment 1.

As illustrated in (1–2), both types of relative-clause sentences involve a nonadjacent dependency between the head-noun, *reporter*, and the main verb, *admitted*, from across the embedded clause. However, ORs additionally involve a backwards-tracking nonlocal dependency (between the embedded verb, *attacked*, and its antecedent object, *reporter*), which generally makes this type of structure more complex. Differential processing difficulty between ORs and SRs is most acute at the main verb, *admitted*, where protracted reading times (RTs) for ORs are evidenced. Individual differences in the degree of comparative difficulty were first reported by King and Just (1991) and linked to variations in verbal working memory (vWM) as assessed by a reading span task. Specifically, individuals with low vWM span scores (“low-span” participants) were found to have slower RT patterns overall than “high-span” participants, as well as a greater divergence in their processing patterns for the two clause types at the main verb region. Within the capacity-based view, the low-span individuals’ poorer processing of ORs has been attributed to limitations in memory resources (e.g., Just and Carpenter, 1992; see also Waters and Caplan, 1996), whereas the experience-based view has emphasized lack of experience with the specific dependency relationships in ORs as the main source of processing difficulty (e.g., MacDonald and Christiansen, 2002).

Here, we test the hypothesis that statistical learning plays a crucial underlying role in shaping readers’ experience of the distributional constraints that govern the less frequent and irregular ORs, which in turn facilitates subsequent RTs. If statistical learning is indeed an important mechanism for such processing phenomena and is meaningfully captured by the new AGL-SRT task, then individual differences in nonadjacent statistical learning (as indexed by Experiment 1) should correlate systematically with inter-individual variation in the ability to track the nonlocal dependency structure of OR sentences. Experiment 2 thus aims to empirically test the strength of this predicted relationship.

METHOD

Participants

The same 50 participants from Experiment 1 provided their consent to participate directly afterwards in this experiment.

Materials

Two experimental sentence lists were prepared, each incorporating 12 initial practice items, 40 experimental items (20 SRs, 20 ORs), and 48 filler items. Yes/No comprehension probes accompanied each sentence item. The SR/OR sentence pairs were taken from Wells et al. (2009) and counterbalanced across the two lists. Semantic plausibility information for subject/object nouns was controlled in the experimental materials.

Procedure

Each participant was randomly assigned to a sentence list, whose items were presented in random order using a standard word-by-word, moving-window paradigm for self-paced reading (Just et al., 1982). Millisecond RTs for each sentence-word and accuracy for each following comprehension question were recorded.

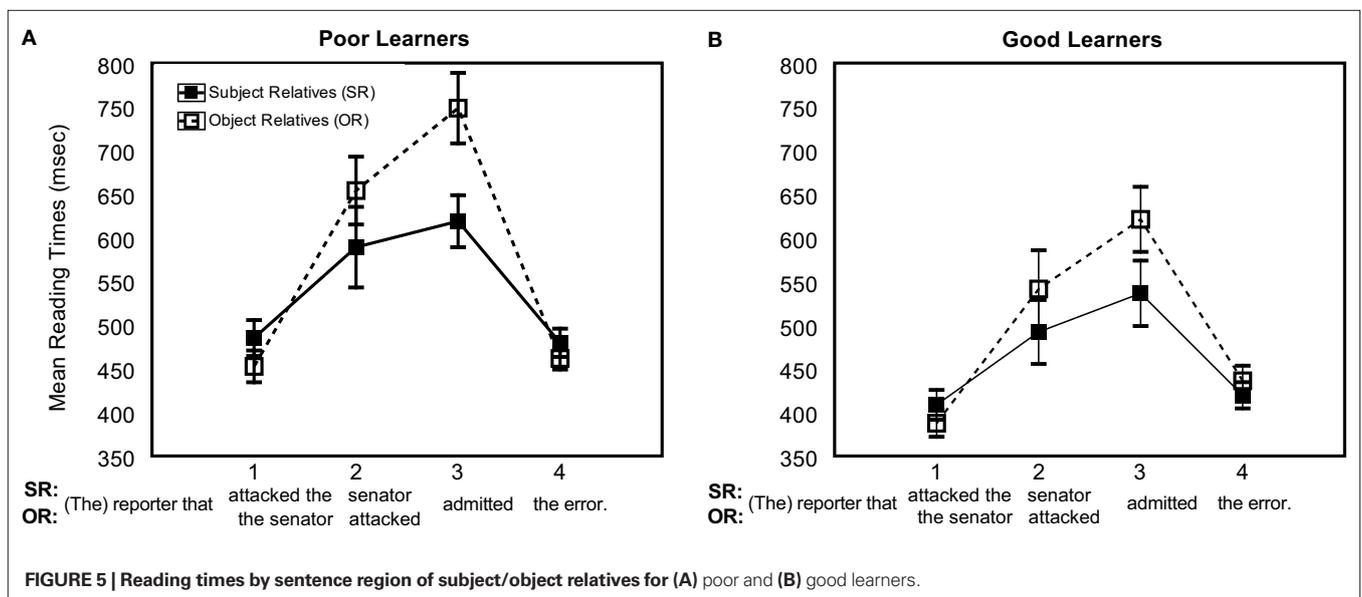
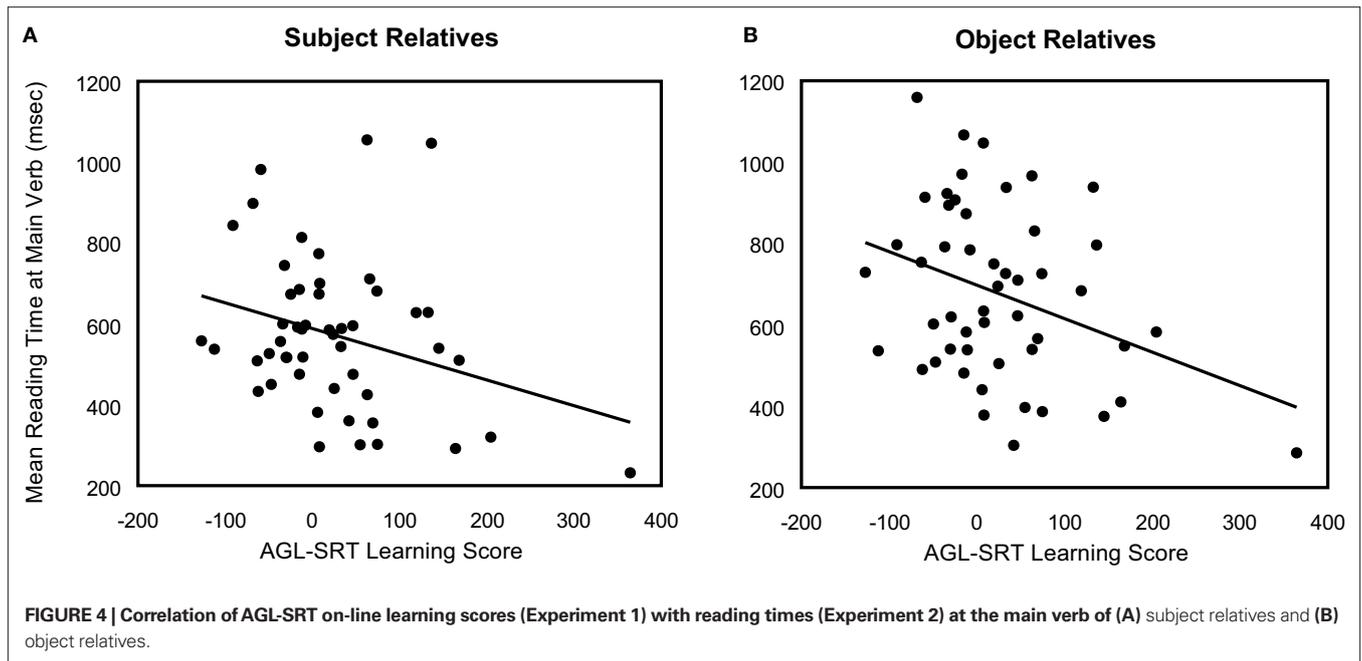
RESULTS AND DISCUSSION

Overall comprehension rate was high ($M = 88.8\%$, $SD = 6.5$). Consistent with past studies, comprehension was lower for ORs ($M = 80.5\%$, $SD = 15.0$) compared to SRs ($M = 87.2\%$, $SD = 9.5$). Data from practice items and incorrectly comprehended sentences were removed from analyses, as were RTs in excess of 2500 ms (0.84% of data). A program error resulted in four participants viewing a small number (8, 4, 2, and 1, respectively) of the experimental/filler items before restarting with a new, complete randomization of their lists’ items. For these participants, only data collected after the restart was used in analyses; however, excluding their data does not alter the pattern of reported results.

Reading times were calculated for the same sentence regions as used in Wells et al. (2009) and prior related work. To test the involvement of statistical learning in mediating individual differences in corresponding language RT patterns, we first assessed correlations between AGL-SRT on-line learning scores (from Experiment 1) and SR/OR main verb RTs. As shown in **Figure 4**, better nonadjacency learning was associated with faster processing for both SRs ($r = -0.30$, $p = 0.04$) and ORs ($r = -0.34$, $p = 0.02$). Across all participants then, individual differences in nonadjacency learning were negatively correlated with grammatical processing difficulty for the relative-clause sentences. Next, for ease of comparison with past work on individual differences in relative clauses processing, we used the same split of good/poor nonadjacency learners from Experiment 1 to compare relative-clause reading patterns of the two groups, as depicted in **Figure 5**.

Compared to poor AGL-SRT learners, good learners tended to exhibit faster RTs at most sentence regions of each clause type, including a significantly quicker mean RT (618.8 ms vs. 748.3 ms) at the critical main verb region of ORs ($F(1,48) = 4.76$, $p = 0.03$). Good AGL-SRT learners also read nominally quicker at the main verb of SRs (536.3 vs. 621.0 ms), but this difference did not reach significance ($F(1,48) = 2.42$, $p = 0.13$). Within groups, poor learners encountered greater difficulty in processing ORs relative to SRs at the main verb, whereas the magnitude of this performance differential was substantially smaller for good learners. The significant group difference for ORs, but nonsignificant difference for SRs, suggests a stronger role for nonadjacent statistical learning skills in the processing of sentences with OR embedded clauses. This is in line with findings from MacDonald and Christiansen (2002), indicating that efficient OR processing may be dependent upon direct experience with the unique dependency structure of ORs, whereas the processing of SRs may benefit from prior experience with the overlapping structure of simple transitive sentences.

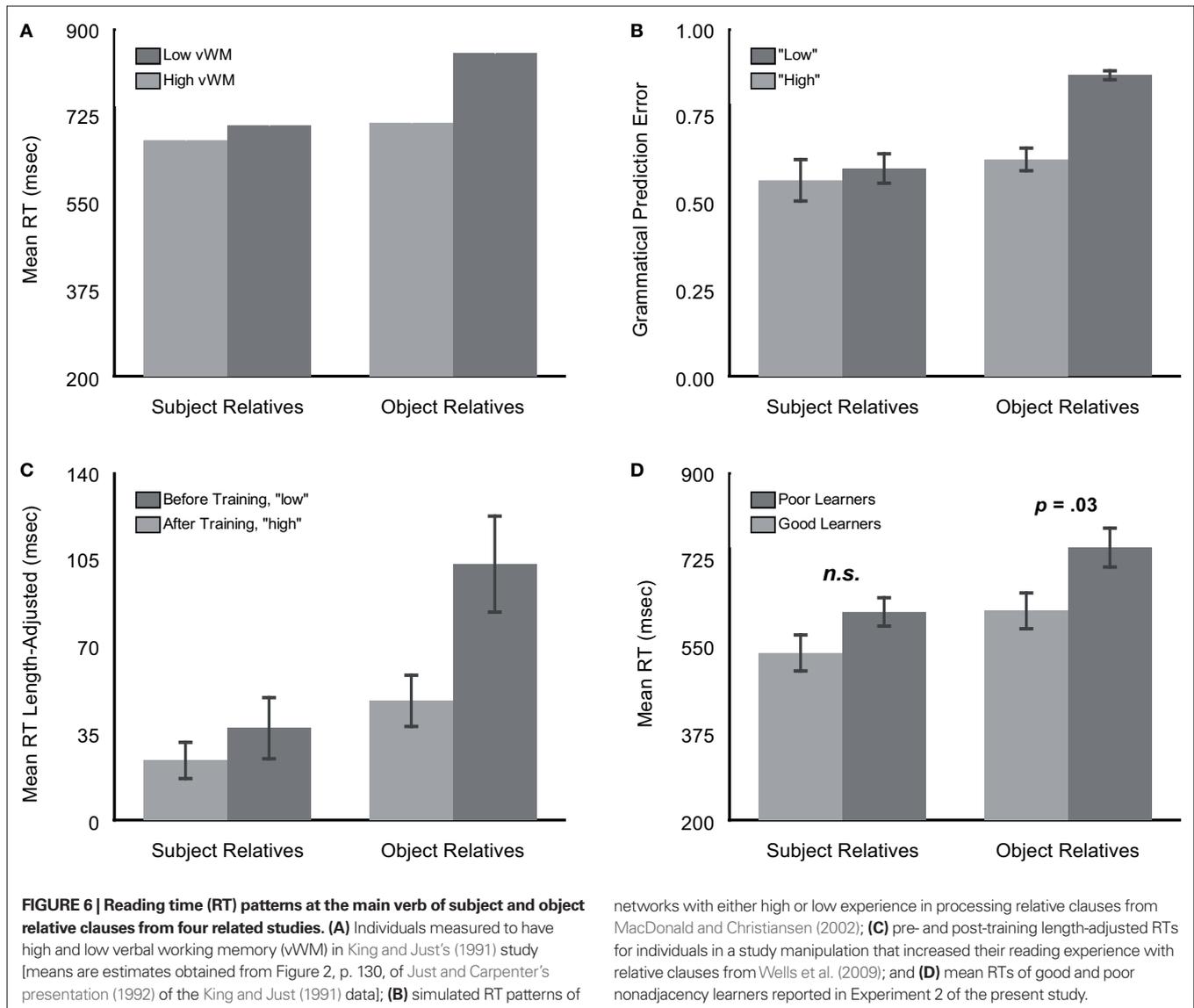
While our AGL-SRT groups differed in their processing of relative clauses, they did not differ on other relevant aspects of AGL-SRT performance, or on the standard grammaticality judgment test from Experiment 1. As noted previously when discussing Experiment 1’s results, individual differences in on-line nonadjacency learning scores (i.e., our statistical learning index) did not correlate with performances on the off-line grammaticality test. This lack of association is present as well when comparing the mean test accuracy of the two AGL-SRT groups, which do not differ (good learners: $M = 60.7\%$, $SE = 3.5$; poor learners: $M = 54.9\%$, $SE = 3.1$; $F(1,48) = 1.42$, $p = 0.24$). There were also no differences between groups on the proportion of errors made across the



training blocks of the AGL-SRT task, over which the on-line learning score was calculated. That is, there were no group differences in the accuracy with which participants selected the appropriate three targets across a string trial (good learners: $M = 91.0\%$, $SD = 6.5$; poor learners: $M = 91.8\%$, $SD = 4.6$; $F(1,48) = 0.20$, $p = 0.66$) nor in the groups' selection accuracy for the final string-element (good learners: $M = 96.3\%$, $SD = 3.2$; poor learners: $M = 96.4\%$, $SD = 2.3$; $F(1,48) = 0.02$, $p = 0.89$). Thus, both good and poor learners were equally engaged and alert in the AGL-SRT task.

Figure 6 places these sentence processing differences within the context of previously observed findings in the literature. Panel A depicts RT patterns for individuals measured to have "high" and "low" vWM in the original King and Just (1991) study. Whereas

high-span individuals were observed to have generally faster RTs, relative performance difficulty at the main verb of object relatives was most pronounced for low-span individuals. Shown in panel B, MacDonald and Christiansen (2002) qualitatively fit the aforementioned RT patterns as a function of the amount of "low" or "high" exposure to relative clauses received by simple recurrent networks (SRNs). Wells et al. (2009) further conducted a human training study whereby they found that increased reading exposure to relative clauses altered participants' RT patterns towards resembling those of the aforementioned "high"-span individuals (and "high" trained SRNs), as illustrated in panel C. The performance contrast between poor and good statistical learners observed in our study (panel D) thus closely mirrors these previous RT patterns



documented in the literature. Namely, the two trends evident in our study (viz., for good contra poor learners: faster overall RTs, and less comparative processing difficulty at the main verb of ORs) reproduce the signature reading patterns documented in the literature for those characterized as having “high” versus “low” vWM span scores respectively. These findings suggest that skill in learning and applying statistical knowledge of distributional regularities, as indexed by on-line learning scores from the novel AGL-SRT paradigm, is involved in natural language processing of relative clauses.

GENERAL DISCUSSION

Although it is typically assumed that statistical learning taps into the same mechanisms that also subserve language (e.g., Gómez and Gerken, 2000; Saffran, 2003), few studies have tested this relationship directly using a within-subjects design to determine whether individuals who are good at statistical learning are also

good at language. We therefore investigated individual differences in statistical learning using a novel AGL-SRT paradigm in Experiment 1, revealing considerable variation between participants in both the on-line trajectory across training and the outcome of learning. Experiment 2 showed that such differences in on-line nonadjacency learning varied systematically with the on-line processing of long-distance dependencies in OR sentences. Together, these results suggest that individual differences in statistical learning are associated with inter-individual variation in language processing and, further, are consistent with the assumption that statistical learning and language may involve the same neurocognitive mechanisms (see also, Petersson et al., 2004).

Given the correlational nature of our design, our study possesses two main limitations: it cannot prove causality, and its observations could in principle be affected by other additional underlying factors, such as participants' eagerness to participate in our experiments. However, given that our groups did not differ on the off-line

AGL-SRT test, it seems less likely that a common method bias could explain the key results. Moreover, as the participants did not differ in errors across training (which were not at floor levels), the lack of differences across these two dimensions also suggests that alertness or task-engagement did not play an important role in the observed pattern of findings. Hence, while a correlational study such as ours cannot explain why a relationship exists, this study does implicate a role in language processing for a largely neglected source of variation: differences in the ability to learn from experience through statistical learning mechanisms. By documenting an association between on-line nonadjacency learning and sentence processing, it provides evidence for a link between the two that future studies may investigate in more detail.

In terms of methodological innovation, the AGL-SRT paradigm in the current study cross-modally instantiated an *artificial language* within an adapted SRT format. This allowed us to take advantage of the structural complexity afforded by artificial languages while at the same time being able to obtain an on-line measure of learning. An analogous prior instantiation of the SRT method with artificial grammar strings, but all visual stimuli, has been employed by Hunt and Aslin (2010). Distinct from other similar approaches (e.g., Cleeremans and McClelland, 1991; Hunt and Aslin, 2001; Howard et al., 2008; Remillard, 2008), however, our paradigm specifically endeavored to capture the continuous timecourse of statistical processing, rather than contrasting/altering the forms of statistical information. Moreover, the AGL-SRT task was designed for the briefer periods of exposure typically associated with statistical learning studies and was applied towards investigating nonadjacency learning as it unfolded on-line.

On the theoretical side, our results are consistent with the idea that the effect of linguistic experience on relative clause processing may be substantially mediated by mechanisms for statistical learning. But what kind of mechanism might be able to accommodate such learning? In a separate study, we found that association-based learning may provide an appropriate candidate mechanism to capture the results of the current findings (Misyak et al., 2010). Computational simulations showed that SRNs were able to capture not only the mean performance trajectory of humans in the new AGL-SRT paradigm, but also the full and wide range of individuals' corresponding off-line scores, without any manipulation of memory-related parameters. These computational results dovetail with previous SRN simulations that reproduced the SR/OR reading

time patterns of low- and high-span individuals as a function of the amount of training exposure, again without manipulating memory (MacDonald and Christiansen, 2002). The SRN simulations further predicted that ORs should be differentially affected by increased exposure to relative-clause sentences, independently of working memory capacity. Wells et al. (2009) empirically confirmed this prediction in their human training study, showing that greater SR/OR reading experience (compared to that of a control condition) tuned reading time profiles towards resembling those of high-span individuals and qualitatively fit the performance of the aforementioned SRNs after the most training exposure. Taken together, the two sets of simulations and the results of the human training study thus argue against the idea that variations in working memory capacity may explain the individual-differences results in Experiments 1 and 2. Moreover, previous studies of individual differences in incidental sequence learning have found no effects of vWM, digit span, or IQ on SRT performance (Feldman et al., 1995; Unsworth and Engle, 2005). This accords with other findings that statistical learning is a better predictor of language processing skills than vWM, cognitive motivation, and nonverbal IQ (Misyak and Christiansen, in press; see also Conway et al., 2010). Results from Misyak and Christiansen (in press) also suggest that statistical learning of adjacent and nonadjacent dependencies are each positively associated with comprehension abilities for different types of natural language sentences, respectively.

To conclude, although the main findings are correlational and thus cannot establish a direct causal link, they nonetheless provide encouraging evidence suggesting that variations in statistical learning ability may be a key factor in mediating the impact of linguistic experience on individual differences in language processing. It is conceivable that variations in statistical learning across individuals may, in turn, be partially determined biologically. Indeed, differences in SRT performance among adolescents with and without SLI have been associated with differential grammatical abilities (Tomblin et al., 2007). Future work studying individual differences in normal and impaired populations using sensitive, on-line measures, as here, will be needed to further elucidate the interrelationships between statistical learning and language.

ACKNOWLEDGMENTS

We thank Christopher Conway, James Cutting, and Rick Dale for comments on a previous version of this paper.

REFERENCES

- Baldwin, D., Andersson, A., Saffran, J., and Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition* 106, 1382–1407.
- Bates, E., Dale, P. S., and Thal, D. (1995). "Individual differences and their implications for theories of language development," in *Handbook of Child Language*, eds P. Fletcher and B. MacWhinney (Oxford: Basil Blackwell), 96–151.
- Cherry, K. E., and Stadler, M. E. (1995). Implicit learning of a nonverbal sequence in younger and older adults. *Psychol. Aging* 10, 379–394.
- Cleeremans, A., Destrebecqz, A., and Boyer, A. (1998). Implicit learning: News from the front. *Trends Cogn. Sci.* 2, 406–416.
- Cleeremans, A., and McClelland, J. L. (1991). Learning the structure of event sequences. *J. Exp. Psychol. Gen.* 120, 235–253.
- Cohen, A., Ivry, R. I., and Keele, S. W. (1990). Attention and structure in sequence learning. *J. Exp. Psychol. Learn. Memory, Cogn.* 16, 17–30.
- Conway, C. M., Bauernschmidt, A., Huang, S. S., and Pisoni, D. B. (2010). Implicit statistical learning in language processing: word predictability is the key. *Cognition* 114, 356–371.
- Destrebecqz, A., and Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychon. Bull. Rev.* 8, 343–350.
- Feldman, J., Kerr, B., and Streissguth, A. P. (1995). Correlational analyses of procedural and declarative learning performance. *Intelligence* 20, 87–114.
- Fiser, J., and Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape-sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 130, 658–680.
- Fiser, J., and Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proc. Natl. Acad. Sci.* 99, 15822–15826.
- Gebhart, A. L., Newport, E. L., and Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychon. Bull. Rev.* 16, 486–490.
- Gómez, R. (2002). Variability and detection of invariant structure. *Psychol. Sci.* 13, 431–436.
- Gómez, R. L., and Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends Cogn. Sci.* 4, 178–186.
- Hasher, L., and Zacks, R. T. (1984). Automatic processing of fundamental

- information: the case of frequency of occurrence. *Am. Psychol.* 39, 1372–1388.
- Howard, J. H. Jr., Howard, D. V., Dennis, N. A., and Kelly, A. J. (2008). Implicit learning of predictive relationships in three-element visual sequences by young and old adults. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 1139–1157.
- Hunt, R. H., and Aslin, R. N. (2001). Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *J. Exp. Psychol. Gen.* 130, 658–680.
- Hunt, R. H., and Aslin, R. N. (2010). Category induction via distributional analysis: evidence from a serial reaction time task. *J. Mem. Lang.* 62, 98–112.
- Jiménez, L., Méndez, C., and Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 948–969.
- Just, M. A., and Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychol. Rev.* 99, 122–149.
- Just, M. A., Carpenter, P. A., and Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *J. Exp. Psychol. Gen.* 111, 228–238.
- King, J., and Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *J. Mem. Lang.* 30, 580–602.
- Kirkham, N. Z., Slemmer, J. A., and Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* 83, B35–B42.
- Lany, J., and Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychol. Sci.* 19, 1247–1252.
- MacDonald, M. C., and Christiansen, M. H. (2002). Reassessing working memory: a comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychol. Rev.* 109, 35–54.
- Misyak, J. B., and Christiansen, M. H. (in press). Statistical learning and language: an individual differences study. *Lang. Learn.*
- Misyak, J. B., Christiansen, M. H., and Tomblin, J. B. (2010). Sequential expectations: the role of prediction-based learning in language. *Top. Cogn. Sci.* 2, 138–153.
- Newport, E. L., and Aslin, R. N. (2004). Learning at a distance I. Statistical learning of nonadjacent dependencies. *Cogn. Psychol.* 48, 127–162.
- Nissen, M. J., and Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cogn. Psychol.* 19, 1–32.
- Onnis, L., Christiansen, M. H., Chater, N., and 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* (Mahwah, NJ: Lawrence Erlbaum Associates), 886–891.
- Pacton, S., and Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 80–96.
- Peterson, K. M., Forkstam, C., and Ingvar, M. (2004). Artificial syntactic violations activate Broca's region. *Cogn. Sci.* 28, 383–407.
- Reali, F., and Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *J. Mem. Lang.* 57, 1–23.
- Reber, A. (1967). Implicit learning of artificial grammars. *J. Verbal Learn. Verbal Behav.* 6, 855–863.
- Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. New York: Oxford University Press.
- Remillard, G. (2008). Implicit learning of second-, third-, and fourth-order adjacent and nonadjacent sequential dependencies. *Q. J. Exp. Psychol.* 61, 400–424.
- Saffran J. R. (2002). Constraints on statistical language learning. *J. Mem. Lang.* 47, 172–196.
- Saffran, J. R. (2003). Statistical language learning: mechanisms and constraints. *Curr. Dir. Psychol. Sci.* 12, 110–114.
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.
- Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996b). Word segmentation: the role of distributional cues. *J. Mem. Lang.* 35, 606–621.
- Thomas, K. M., and Nelson, C. A. (2001). Serial reaction time learning in pre-school- and school-age children. *J. Exp. Child Psychol.* 79, 364–387.
- Tomblin, J. B., Mainela-Arnold, E., and Zhang, X. (2007). Procedural learning in adolescents with and without specific language impairment. *Lang. Learn. Dev.* 3, 269–293.
- Unsworth, N., and Engle, R. W. (2005). Individual differences in working memory capacity and learning: evidence from the serial reaction time task. *Mem. Cogn.* 33, 213–220.
- Waters, G. S., and Caplan, D. (1996). The capacity theory of sentence comprehension: critique of Just and Carpenter (1992). *Psychol. Rev.* 103, 761–772.
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., and MacDonald, M. C. (2009). Experience and sentence processing: statistical learning and relative clause comprehension. *Cogn. Psychol.* 58, 250–271.
- Willingham, D. B., Nissen, M. J., and Bullemer, P. (1989). On the development of procedural knowledge. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 1047–1060.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 10 March 2010; paper pending published: 03 May 2010; accepted: 29 June 2010; published online: 14 September 2010.
 Citation: Misyak JB, Christiansen MH and Tomblin JB (2010) On-line individual differences in statistical learning predict language processing. *Front. Psychology* 1:31. doi: 10.3389/fpsyg.2010.00031
 This article was submitted to *Frontiers in Language Sciences*, a specialty of *Frontiers in Psychology*.
 Copyright © 2010 Misyak, Christiansen and Tomblin. This is an open-access article subject to an exclusive license agreement between the authors and the *Frontiers Research Foundation*, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.