Is statistical learning trainable?

Luca Onnis (lucao@ntu.edu.sg)
Matthew Lou-Magnuson (matt0020@ntu.edu.sg)
Division of Linguistics and Multilingual Studies, Nanyang Technological University
Singapore, 637332

Hongoak Yun (oaktreepark@gmail.com)
Department of Media and Communication, Konkuk University
Seoul, Korea, 143-701

Erik Thiessen (thiessen@andrew.cmu.edu)
Department of Psychology, Carnegie Mellon University
Pittsburgh, PA, 15213

Abstract

Statistical learning (SL) is the ability to implicitly extract regularities in the environment, and likely supports various higher-order behaviors, from language to music and vision. While specific patterns experience are likely to influence SL outcomes, this ability is tacitly conceptualized as a fixed construct, and few studies to date have investigated how experience may shape statistical learning.

We report one experiment that directly tested whether SL can be modulated by previous experience. We used a pre-post treatment design allowing us to pinpoint what specific aspects of "previous experience" matter for SL. The results show that performance on an artificial grammar learning task at post-test depends on whether the grammar to be learned at post-test matches the underlying grammar structures learned during treatment. Our study is the first to adopt a pre-post test design to directly modulate the effects of learning on learning itself.

Keywords: experience; implicit learning; pre-post test design; sequential learning; statistical learning; statistical training; transition probabilities.

Introduction

Statistical learning (SL) is the ability to acquire patterned regularities in the environment, and abstract over them to find structural relations among sequences of stimuli and events. This core human ability is believed to involve a set of basic cognitive mechanisms that are sensitive to probabilistically distributed spatial and sequential information.

Underlying most of the research in SL to date is the tacit assumption that SL mechanisms are a fixed set, and do not change substantially in individuals as a function of either time or experience (Reber, 1993). This assumption is corroborated by studies finding parallel results when infants, children, and adults are exposed to similar statistical cues and similar tasks, reinforcing the ubiquitous role of SL in human cognition across the lifespan. For example, Meulemans, Van der Linden, and Perruchet (1998) found no age-related performance differences in a serial reaction time in 6- or 10-year-olds, nor in adults, pointing to the stable persistence of this ability in humans (although see Simon et al., 2011 for evidence of a decline in ageing population). Here we challenge the long-held assumption of fixedness of SL abilities, and ask whether SL can be augmented through prior learning experiences. In particular we ask whether participants in tasks that involve tracking probabilistic relations among sequential stimuli can be trained to implicitly attend to statistics that differ from the ones they would be normally inclined to detect.

If statistical learning involves a form of optimization and reduction of uncertainty about the environment (Onnis, Christiansen, Chater, & Gomez, 2003), we would expect statistical learning to result in adaptations to the specific sensory environments of the learner. A few studies have begun to assess this possibility of *learning to learn*. In most of these cases, however, the adaptations involved characteristics of the input (such as phonotactic or phonological patterns) that might be learned via mechanisms other than statistical learning (e.g., Saffran & Thiessen, 2003). Lany and Gomez (2008) found that in 12-month olds the initial familiarization with adjacent dependencies resulted in enhanced learning of nonadjacent dependencies, a type of relation that is difficult to learn at that age. Similarly, word learning in infants (Graf-Estes et al., 2007) and adults (Mirman et al., 2008) can be facilitated when words are consistent with expectations based on a previous word segmentation task containing statistical cues to word boundary.

These studies established that learners can build on certain types of computations to better attend to *other* regularities in the input for different aspects of language learning. However, they did not assess directly whether the *same* statistical computations that learners process by default are altered by their past experience.

One recent experiment, though, found that the patterns of transitional probabilities to which learners are sensitive varies as a function of linguistic background (Onnis & Thiessen, 2013). The authors interpreted these results to mean that years of exposure to specific statistical patterns in different linguistic environments induced changes in statistical learning of novel artificial grammars. This interpretation suggests that learners carry with them statistical biases developed over years of exposure to their native language(s) that lead to different expectations about novel subsequent input. Onnis & Thiessen (2013) represents the point of departure of the current study. Because a putative correlation between language background and subsequent performance in a SL task cannot establish any direction of causality, in this study we set out to directly manipulate the statistical landscape of a first artificial grammar to establish whether it differentially affects parsing preferences in a subsequent artificial grammar task. To our knowledge, our pre-post test design is the first to be applied to the statistical learning literature to assess effects of experience on learning.

Method

We exposed participants consecutively to two artificial grammars that shared no surface features, and looked at participants' preferences for grouping percepts – here syllables – based on how they utilized information from forward and backward transition probabilities (TPs). A pre-post test design manipulated two statistical training conditions, allowing us to directly assess their impact on subsequent statistical learning.

Participants. We recruited 62 Korean students at Konkuk University, Seoul. They received the equivalent of US$9 for their participation.

Materials. Two types of artificial grammar were created. Grammar A was a sequence of 711 letter symbols generated according to the rules of a stochastic Markovian grammar chain. The process started by choosing one of eight possible symbols (X, Y, A, B, C, D, E, F) at random, and then generating the next symbol according to a set of probabilistic sequencing rules. The actual sequence was realized as the continuous concatenation of eight monosyllabic words to form a pauseless 3.5 minute speech stream. We randomly assigned each letter placeholder to a given word, in which 80 ms was allotted for consonants and 260 ms for vowels. Because we were interested in the perception of grouping boundaries as driven by statistical biases alone, we synthesized the speech stimuli eliminating possible prosodic cues. The resulting audio was faded in and faded out over 5 seconds, giving the impression of an infinite loop. The Italian diphone set in MBROLA (http://tcts.fpms.ac.be/synthesis/mbrola.html) created words that sounded different in vocal quality from Korean, but were still clearly perceivable, so as to engage participants in an "alien language" learning task. All phonemes had equivalent phonemic realizations in Korean, and all syllable sequences were phonotactically permitted. Crucially, the sequence of concatenated syllables in Grammar A contained *conflicting* forward and backward transitional probabilities, as in the following training sample:

a) .. *fushezirafunizitifugezibu* ..

Previous research has shown that infants and adults alike are independently sensitive to both forward and backward transition probabilities, and use them implicitly to make judgments about the likely groupings of otherwise unsegmented streams of speech. Unlike previous studies in which forward and backward TPs worked *together* to assist speech segmentation (e.g., Saffran et al., 1996), in our Grammar A, forward TPs and backward TPs were pitted against each other and thus competed as cues for grouping boundaries. Thus, Grammar A was statistically ambiguous such that whenever forward probability was low between any two adjacent syllables, (e.g., fwd-TP(*zi*|*she*) = .33), backward probability was high (back-TP(*she*|*zi*) = 1), and vice versa (e.g., fwd-TP(*ra*|*zi*) = 1; back-TP(*zi*|*ra*) = .33). A consequence of this statistical ambiguity is that two parses of sample a) into two-syllable linguistic units were thus equally acceptable and possible – namely parses b) and c) below. In b), one segments the signal such that the two syllables of a unit have a high forward probability and a low backward probability (the Hi-Lo patterns), while in c), the word-internal forward probabilities are low and the backward probabilities are high (the Lo-Hi pattern).

b) (Hi-Lo parse):*..fushe zira funi ziti fuge zibu ..*

c) (Lo-Hi parse): *.. shezi rafu nizi tifu gezi ..*

At test, two-word groupings corresponding to the Hi-Lo and Lo-Hi patterns were pitted one against the other in a two-alternative forced-choice (AFC) task. Six test pair trials were presented in random sequential order, while the order within a pair was counterbalanced by repeating each test pair twice, for a total of 12 test trials.

The other type of grammar was split into two varieties, namely, B-forward and B-backward, and their sequential structure contained opposite statistical cues to group boundaries. We first created a training corpus of 12 new synthesized monosyllabic words ('do' 'te' 'ma' 'ke' 'ne' 'tu' 'bi' 'ge' 'da' 'pa' 'vo' 'po'). The total length of the sequence, as well as the length of individual syllables, were the same as Grammar A, with the difference that the voice used was male. Likewise, test stimuli were created by grouping two monosyllabic words at a time at each transition point in the stream.

Each of these two Grammar B varieties were randomly assigned to participants. Modeled after the grammar of Perruchet & Desaulty (2008; see also Onnis & Thiessen, 2013), in the Grammar B-Forward, forward transition probabilities were informative, with syllables alternating between high and low forward TPs (1 and 0.11 respectively). Backward probabilities were uninformative. In Grammar B-Backward ,backward transition probabilities were informative, with words alternating between high and low backward TPs (1 and 0.11 respectively). This time forward probabilities were uninformative.

We reasoned that participants may perceive grouping boundaries in the sequence probabilistically when it is most informative (i.e., whenever the backward or forward TPs are lowest) in accordance with previous research (Perruchet & Desaulty, 2008; Pelucchi et al., 2009). Thus, we expected the two Grammar Bs to differentially bias learners toward one type of transition probability, resulting in different groupings at test.

Procedure. All participants listened to the same Grammar A at pre- and post-test, while they were randomly assigned to one of the Grammar Bs during treatment/biasing. Listening lasted 3.5 min for both grammars. The forced-choice tests occurred at the end of each listening session. As customary for these types of segmentation experiments, for each stimuli pair participants were asked to choose which one formed a grouping in the sequence they had just heard. Participants wore headphones and instructions were administered in Korean. The experiment was conducted on two separate days one week apart. Pre-test (Grammar A) on Day 1, and treatment + post-test (Grammar B + Grammar A again) on Day 2.

Results

We analyzed results separately in terms of group scores at pre-test, treatment, and in terms of change of scores between pre- and post-test. In all analyses and for each test trial LoHi preference was coded as 1 and Hilo preference was coded as 0. We then averaged across test trials to obtain a proportion preference for LoHi patterns for each participant.

Pretest results. After establishing that the distribution of participants' preference for LoHi patterns was not uniform (Shapiro-Wilk normality test, $W = 0.96$, $p = 0.026$), the nonparametric one-way Wilcoxon signed-rank test with continuity correction revealed that participants' mean as a group was not significantly different from the random baseline of 0.5 (M = 0.52, SD = 0.19, V = 746.5, p = 0.70). As apparent from the standard deviation, there was individual variation in participants' degree of bias. Further inspection revealed a non-unimodal distribution of data, suggesting one mode with preference for HiLo patterns and another one with preference for LoHi patterns. This may be due to differences in participants' linguistic backgrounds, as many of the participants spoke multiple languages, which may influence their bias for forward-going or backward-going directionality (Onnis & Thiessen, 2013). Regardless, these results demonstrate that as a group, these participants have no strong bias for forward-going or backward-going TPs.

Treatment results. Participants assigned to Grammar B-backward preferred LoHi patterns (*M = 0.73*, *SD = 0.18*), while those assigned to Grammar B-forward dispreferred LoHi patterns (*M = 0.37*, *SD = 0.23*), showing a preference for HiLo patterns. A one-way ANOVA with Biasing condition as a two-level between-subject factor (Grammar B-backward versus Grammar B-forward) revealed highly significant differences between the two Biasing groups, $F(1,59)=41.64$, $p<0.001$). Thus both groups differed in pattern endorsements in opposite directions, and in the direction consistent with the statistical cues present in each grammar. This result suggests that learners attend to different types of statistical cues when they are weighted in such a way to be uniquely informative.

Pre-post test. We noted earlier that some participants demonstrated strong pre-test biases in both directions. To account for this heterogeneity in the pre-test data we computed pre-post test difference scores for each participant as the dependent measure. A one-way ANOVA with Biasing Condition (Grammar B-forward versus Grammar B-backward) as a between-subject factor revealed significant differences between the two Biasing groups, $F(1,58)=5.49$, $p<0.05$ (one participant did not complete the post-test and her difference scores were not computed). Importantly, the differences were in the expected direction (see Figure 1). The Grammar B-forward group did not modify their preference much from pre- to post-test, arguably because the cues for segmenting at low points of forward TP were present in both Grammar A and Grammar B-forward. In comparison, Grammar B-backward contained all cues in the opposite direction, and participants trained on Grammar B-backward shifted (as a group) their initial preference at pre-test upward on the gradient from HiLo to LoHi at post-test. These results suggest that it is possible to alter participants' biases to attend to different statistics in the environment.

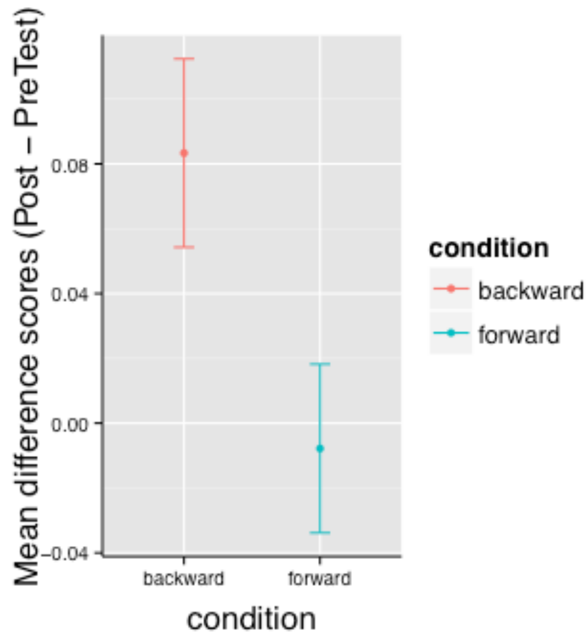

Figure 1. Mean score differences and standard deviations of participants preferences for LoHi patterns, presented by biasing condition. Pre-test scores were subtracted from Post-test scores.

Discussion

Recent research has begun to investigate the effects of experience on statistical learning mechanisms (e.g., Lany & Gomez, 2008; Graf-Estes et al., 2007; among others). These studies indicate that learners can use statistical information in one task to discover different relations in a new task. Our goal was to explore the possibility to modify learners' sensitivity to the *same* statistical relations (here forward and backward probabilities) by directly manipulating and comparing different types of statistical exposure. Trainability of statistical learning was tested with a group of adult Korean speakers who participated in a pre-test, treatment, post-test design. To our knowledge, this is the first application of the design in the statistical learning literature, and has both theoretical and practical implications.

Theoretically, our results challenge the tacit assumption in the literature that implicit SL is a rather fixed skill, and there is relatively small variation, both across (Reber, 1989), and within individuals over time, as well as a function of experience. The possibility that experience with the perceptual world may modify statistical learning proclivities has been recently explored in the realm of language (e.g., Onnis & Thiessen, 2013) and music (Shook et al., 2013). While those studies hinted at possible correlations between SL and experience, our pre-post test design helped establish an initial direct causal relation between specific types of statistical experience and their subsequent impact on learners' statistical biases.

Our study also explored the possibility of transfer for trained statistical bias to novel stimuli. Importantly, such transfer of statistical penchant can occur in the absence of surface similarity between the stimuli – here Grammar A and B did *not* share the same syllables. This is an important finding, because detractors of SL often see it as a set of mechanisms that operates only on the basis of surface similarity. Thus, SL may occur as a result of more powerful abstractive processes.

Our results invite future explorations into individual differences in statistical learning. Heterogeneity in both pre- and post-tests indicate that some learners had stronger initial biases, and/or they were be more or less susceptible to the biasing conditions than others. This has implications for studies that directly relate the ability to learn statistically to the ability to learn and process language. Infants exhibit individual differences in statistical learning skills that may modulate language development trajectories (e.g., Kidd, 2012). Direct predictive relations between statistical learning scores and online sentence processing and other linguistic tasks exist now both for children and adults (Yim & Windsor, 2010). In addition, neurophysiological studies using within-subject designs suggest that similar neural mechanisms serve both syntactic processing of language and statistical learning of sequential patterns (Abla, Katahira, & Okanoya, 2008; Christiansen, Conway, & Onnis, 2007). Our results using a pre-post design suggest the applicability of future statistical training regimes adapted for non-normally developing populations of language learners. Studies with children have now linked poor implicit and statistical learning skills with reading difficulties (Yim & Windsor, 2010), developmental dyslexia (Hedenius et al., 2013), and specific language impairments (Hsu, Tomblin, & Christiansen, 2014; Lukács & Kemény, 2014). Poor language learners may be less sensitive to the statistics inherent in natural language, or may pick up statistics of the language that are less relevant to discovering linguistic structure. If this were so, one attractive potential of statistical learning training would be to help learners implicitly optimize their learning process, by targeting and scaffolding statistical relations that are harder for them to internalize.

While the effect of experience we saw is in line with our predictions, it is preliminary and future research and comparisons would benefit from a larger effect size, which may be possible with alterations to this paradigm. First, exposure to the Grammar B's was limited to 4 minutes. Perhaps longer exposure would make the bias more robust. (though note that the post-test suggests participants picked up the bias, suggesting that increasing exposure may have limited effect) Another possible explanation for the small effect size of training is that most participants reported remembering Grammar A at post-test from the previous week at pre-test. Thus, the desired bias may have been obfuscated by participants' fresh memories of the first exposure to the grammar at pre-test. A better way to promote transfer would thus be to create a novel Grammar C at post-test that has the same underlying sequence as Grammar A but no feature resemblance to either Grammar A and the Grammar B's.

Another way to improve the paradigm would involve exposing participants to a bias in more modalities or stimulus sets. Generally speaking, experiencing a pattern in multiple different contexts makes that pattern more generalizable to novel contexts (e.g, Hintzman, 1986; Lively, Logan, & Pisoni, 1993). Thus, participants could be exposed to forward or backward patterns in shapes and tones, or perhaps instantiated by multiple different speakers (as opposed to one speaker) to promote generalization and transfer.

In conclusion, we have shown that participants' preference to parse a string of phonemes according to forward or backward transition probabilities can be directly manipulated via prior statistical learning experiences. Thus, rather than being a fixed process that learners use, implicit statistical learning might itself be a process that is trainable.

Acknowledgments

This study was supported by Nanyang Technological University's Start-Up-Fund grant #M4081274 to the first author.

References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience*, *20*(6), 952-964.

Christiansen, M. H., Conway, C. M., & Onnis, L. (2012). Similar neural correlates for language and sequential learning: evidence from event-related brain potentials. *Language and cognitive processes*, *27*(2), 231-256.

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18(3)*, 254-260.

Hedenius, M., Persson, J., Alm, P. A., Ullman, M. T., Howard, J. H., Howard, D. V., & Jennische, M. (2013). Impaired implicit sequence learning in children with developmental dyslexia. *Research in developmental disabilities*, *34*(11), 3924-3935.

Hintzman, D.L. (1986). "Schema Abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411-428.

Hsu, H. J., Tomblin, J. B., & Christiansen, M. H. (2013). Impaired statistical learning of non-adjacent dependencies in adolescents with specific language impairment. *Frontiers in psychology*, *5*, 175-175.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition, 116(3)*, 321-340.

Kidd, E. (2012). Implicit statistical learning Is directly associated with the acquisition of syntax. *Developmental psychology, 48(1)*, 171-184.

Lively, S.E., Logan, J.S., & Pisoni, D.B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*, 2962-2973.

Lukács, Á., & Kemény, F. (2014). Domain-general sequence learning deficit in specific language impairment. *Neuropsychology*, *28*(3), 472-483.

Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, *108(1)*, 271-2

Meulemans, T., Van der Linden, M., & Perruchet, P. (1998). Implicit sequence learning in children. *Journal of experimental child psychology*, *69*(3), 199-221.

Moreau, D., & Conway, A. R. (2014). The case for an ecological approach to cognitive training. *Trends in cognitive sciences*, *18*(7), 334-336.

Oei, A.C. and Patterson, M.D. (2013) Enhancing cognition with video games: a multiple game training study. *PLoS ONE 8*, e58546.

Onnis, L., Christiansen, M. H., Chater, N., & Gómez, R. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial grammar learning. In *Proceedings of the 25th annual conference of the cognitive science society* (pp. 886-891). Mahwah, NJ: Lawrence Erlbaum.

Onnis, L., & Thiessen, E. (2013). Language experience changes subsequent learning. *Cognition*, *126*(2), 268-284.

Shook A, Marian V, Bartolotti J, Schroeder SR. (2013). Musical experience influences statistical learning of a novel language. *American Journal of Psychology*, *126*(1): 95-104.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244-247.

Perruchet, P., & Desaulty, S. (2008). A role for backward transitional probabilities in word segmentation?. *Memory & Cognition*, *36*(7), 1299-1305.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219–235.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental psychology*, 39(3), 484-494.

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

Simon, J. R., Howard, J. H., & Howard, D. V. (2011). Age differences in implicit learning of probabilistic unstructured sequences. T*he Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *66*(1), 32-38.

Yim, D., & Windsor, J. (2010). The roles of nonlinguistic statistical learning and memory in language skill. *Korean Journal of Communication Disorders*, *15*, 381-396.