Title
Statistical Learning Within and Across Modalities: Abstract versus Stimulus-Specific Representations

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 27(27)

ISSN
1069-7977

Authors
Christiansen, Morten H.
Conway, Christiopher T.

Publication Date
2005

Peer reviewed

Statistical Learning Within and Across Modalities: Abstract versus Stimulus-Specific Representations

Christopher M. Conway (cmc82@cornell.edu) **Morten H. Christiansen (mhc27@cornell.edu)**
Department of Psychology, Uris Hall, Cornell University
Ithaca, NY 14853 USA

Abstract

When learners encode sequential patterns and generalize their knowledge to novel instances, are they relying on abstract or stimulus-specific representations? Artificial grammar learning (AGL) experiments showing transfer of learning from one stimulus set to another has encouraged the view that learning is mediated by abstract representations that are independent of the sense modality or perceptual features of the stimuli. Using a novel modification of the standard AGL paradigm, we present data to the contrary. These experiments pit abstract, domain-general processing against stimulus-specific learning. The results show that learning in an AGL task is mediated to a greater extent by stimulus-specific, rather than abstract, representations. They furthermore show that learning can proceed separately and independently (i.e., in parallel) for multiple input streams that occur along separate perceptual dimensions or modalities. We conclude that learning probabilistic structure and generalizing to novel stimuli inherently involves learning mechanisms that are closely tied to perceptual features.

Keywords: statistical learning; artificial grammar learning; modality-specificity; crossmodal; intramodal

Introduction

The world is temporally bounded. The events that we observe, as well as the behaviors we produce, occur sequentially over time. It is therefore important for organisms to have the ability to process sequential information. One way of encoding sequential structure is by learning the statistical relationships between sequence elements occurring in an input stream. Statistical learning of sequential structure is involved in many aspects of human and primate cognition, including skill learning, perceptual learning, and language processing (Conway & Christiansen, 2001).

Statistical learning has been demonstrated in many domains, using auditory (Saffran, Johnson, Aslin, & Newport, 1999; Saffran, Newport, & Aslin, 1996), visual (Baker, Olson, & Behrmann, 2004; Fiser & Aslin, 2002), tactile (Conway & Christiansen, 2005), and visuomotor stimuli (Cleeremans & McClelland, 1991; Nissen & Bullemer, 1987). However, several questions remain unanswered. For instance, it is not entirely clear to what extent learning is specific to the input modality in which it is learned. This has been a hotly debated issue in cognitive science (e.g., Christiansen & Curtin, 1999; Marcus, Vijayan, Rao, & Vishton, 1999; McClelland & Plaut, 1999; Seidenberg & Elman, 1999). Is statistical learning stimulus-specific or is it abstract and amodal? The traditional "abstractive" view posits that learning consists of extracting the abstract, amodal rules of the underlying input structure (e.g., Marcus et al., 1999; Reber, 1993). Alternatively, instead of abstract knowledge, participants may be learning the statistical structure of the input sequences in a modality- or feature-specific manner (e.g., Chang & Knowlton, 2004; Conway & Christiansen, 2005).

Another unanswered question is: can people learn different sets of statistical regularities *simultaneously* across and within modalities? The answer to this question will help reveal the nature of the underlying cognitive/neural mechanisms of statistical learning. If people can learn multiple concurrent streams of statistical information independently of one another, it may suggest the existence of multiple, modality-specific mechanisms rather than a single amodal one.

A Modified Artificial Grammar Design

One way to explore these issues is by using the artificial grammar learning (AGL) task. In a standard AGL experiment (Reber, 1967), an artificial grammar is used to generate stimuli that conform to certain rules governing the order that elements can occur within a sequence. After being exposed to a subset of structured sequences under incidental learning conditions, it is participants' task to classify novel stimuli in terms of whether they conform to the rules of the grammar. Participants typically achieve a moderate degree of success despite being unable to verbally express the nature of the rules, leading to the assumption that learning is "implicit". Furthermore, because the task presumably requires learners to extract the probabilistic structure of the sequences, such as element co-occurrences, learning can be regarded as one of computing and encoding statistically-based patterns.

We introduce a novel modification of the AGL paradigm to examine the nature of statistical learning within and across modalities. We used two different finite-state grammars in a cross-over design such that the grammatical test sequences of one grammar were used as the ungrammatical test sequences for the other grammar. In the training phase, each grammar was instantiated in a different sense modality (auditory tones versus color sequences, Experiment 1) or within the same modality along different perceptual "dimensions" (colors versus shapes, Experiment 2A; tones versus nonwords, Experiment 2B) or within the same perceptual dimension (two different shape sets, Experiment 3A; or two different nonword sets, Experiment 3B). At test, all sequences were instantiated in just one of the vocabularies they were trained on (e.g., colors or tones for Experiment 1).

For example, in Experiment 1, participants were exposed to visual sequences of one grammar and auditory sequences from the other grammar. In the test phase, they observed new grammatical sequences from both grammars, half generated from one grammar and half from the other. However, for each participant, all test items were instantiated only visually or only aurally.

This cross-over design allows us to make the following prediction. If participants learn the abstract underlying rules of both grammars, they ought to classify all sequences as equally grammatical (scoring 50%). However, if they learn statistical regularities specific to the sense modality in which they were instantiated, participants ought to classify a sequence as grammatical only if the sense modality and grammar are matched appropriately, in which case the participants should score above chance levels. We also incorporated single-grammar conditions to provide a baseline level for comparison to dual-grammar learning.

Experiment 1: Crossmodal Learning

Experiment 1 assesses crossmodal learning by presenting participants with auditory tone sequences generated from one grammar and visual color sequences generated from a second grammar. We then test participants using novel grammatical stimuli from each grammar that are instantiated in one of the vocabularies only (tones or colors), cross-balanced across participants. If participants learn the underlying statistical regularities of the grammars specific to the sense modality in which they were presented, they ought to classify the novel sequences appropriately. On the other hand, if instead participants are learning the abstract, amodal structure of the sequences, all test sequences will appear equally grammatical, and this should be reflected in their classification performance.

Method

Subjects For Experiment 1, 40 participants (10 in each condition) were recruited for extra credit from Cornell University undergraduate psychology classes.

Materials Two different finite-state grammars, Grammar A and Grammar B (shown in Figure 1), were used to generate two sets of non-overlapping stimuli. Each grammar had 9 grammatical sequences used for the training phase and 10 grammatical sequences used for the test phase, all sequences containing between three and nine elements. As Figure 1 shows, the sequence elements were the letters X, T, M, R, and V. For Experiment 1, each letter was in turn instantiated either as one of five differently colored squares or one of five auditory tones. The five colored squares ranged along a continuum from light blue to green, chosen such that each was perceptually distinct yet similar enough to make a verbal coding strategy difficult. The five tones had frequencies of 210, 245, 286, 333, and 389 Hz. These frequencies were chosen because they neither conform to standard musical notes nor contain standard musical intervals between them (see Conway & Christiansen, 2005). As an example, for one participant, the Grammar A sequence "V-V-M" might be instantiated as two light green stimuli followed by a light blue stimulus, whereas for another participant, this same sequence might be instantiated as two 389 Hz tones followed by a 286 Hz tone.

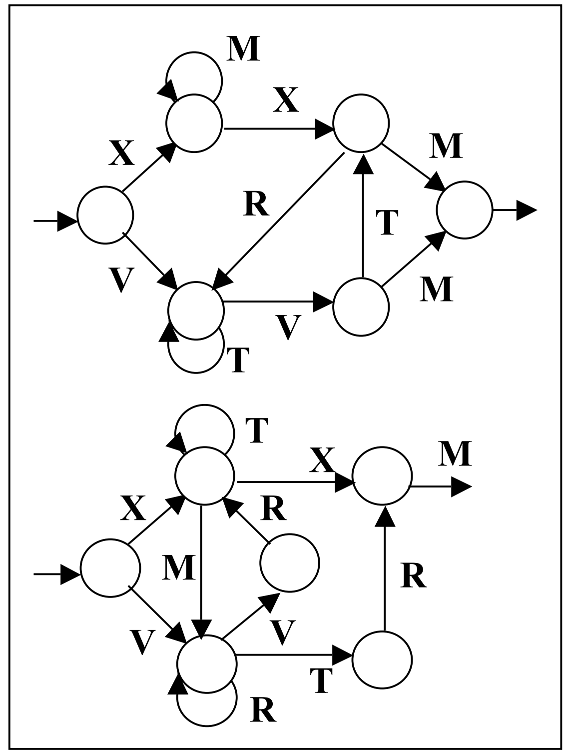

Figure 1: Grammar A (top) and Grammar B (bottom) used in all three experiments. The letters from each grammar were instantiated as colors or tones (Experiment 1), colors or shapes (Experiment 2A), tones or nonwords (Experiment 2B), two different shape sets (Experiment 3A), or two different nonword sets (Experiment 3B).

All visual stimuli were presented in a sequential format in the center of a computer screen. Auditory stimuli were presented via headphones. Each element (color or tone) of a particular sequence was presented for 500ms with 100ms occurring between elements. Each sequence was separated by 1700ms blank screen.

Procedure Participants were randomly assigned to one of two experimental conditions or one of two baseline control conditions. Participants in the experimental conditions were trained on color sequences from one grammar and tone sequences from the other grammar. Modality-grammar assignments were cross-balanced across participants. Additionally, the particular assignment of letters to visual or auditory elements was randomized for each participant. Participants were told that they would hear and/or see sequences of auditory and visual stimuli. Importantly, they were not explicitly told of the existence of the grammars, underlying rules, or regularities of any kind. However, they were told that it was important to pay attention to the stimuli because they would be tested on what they observed. The 18 training sequences (9 from each grammar) were presented randomly, one at a time, in six blocks, for a total of 108 sequences. Note that because the order of presentation was entirely random, the visual and auditory sequences were completely intermixed with one another.

In the test phase, participants were instructed that the stimuli they had observed were generated according to a complex set of rules that determined the order of the stimulus elements within each sequence. Participants were told they would now be exposed to new color or tone sequences that they had not yet observed. Some of these sequences would conform to the same set of rules as before, while the others would be different. Their task was to judge which of the sequences followed the same rules as before and which did not. For the test phase, 20 sequences were used, 10 that were grammatical with respect to one grammar and 10 that were grammatical with respect to the other. For half of the participants, these test sequences were instantiated using the color vocabulary (Visual-Experimental condition), while for the other half, the test sequences were instantiated using the tone vocabulary (Auditory-Experimental condition). A classification judgment was scored as correct if the sequence was correctly classified in relation to the sense modality in question.

Participants in the baseline control conditions followed a similar procedure except that they received training sequences from only one of the grammars, instantiated in just one of the sense modalities, cross-balanced across participants. The nine training sequences were presented randomly in blocks of six for a total of 54 presentations. The baseline participants were tested using the same test set, instantiated with the same vocabulary with which they were trained on. Thus, the baseline condition assesses visual and auditory learning with one grammar alone (Visual-Baseline and Auditory-Baseline conditions).

Results and Discussion

We report mean correct classification scores (out of 20) and t-tests compared to chance levels for each group: 12.7 (63.5%), $t(9)=2.76$, $p<.05$ for the Visual-Experimental condition; 14.1 (70.5%), $t(9)=4.38$, $p<.01$ for the Auditory-Experimental condition; 12.4 (62.0%), $t(9)=2.54$, $p<.05$ for the Visual-Baseline condition; and 13.1 (65.5%), $t(9)=3.44$, $p<.01$ for the Auditory-Baseline condition. Thus, each group's overall performance was better than what would be expected by chance. Furthermore, we compared each experimental group to its respective baseline group and found no statistical differences: Visual-Experimental versus Visual-Baseline, $t(9)=.22$; $p=.83$; Auditory-Experimental versus Auditory-Baseline, $t(9)=1.1$; $p=.30$.

These results clearly show that participants can simultaneously learn statistical regularities from input generated by two separate artificial grammars, each instantiated in a different sense modality[1]. Perhaps surprisingly, the levels of performance in the dual-grammar experimental conditions are no worse than those resulting from exposure to stimuli from just one of the grammars alone. This lack of a learning decrement suggests that learning of visual and auditory statistical structure occurs in parallel and independently. Furthermore, these results stand in contrast to previous reports showing transfer of learning in AGL between two different modalities (e.g., Altmann, Dienes, & Goode, 1995). Our data essentially show a lack of transfer. If our participants had exhibited transfer between the two sense modalities, then all test sequences would have appeared grammatical to them, driving their performance to chance levels. Thus, our data suggests that the knowledge of the statistical patterns, instead of being amodal or abstract, was stimulus-specific. We next ask whether learners can similarly learn from two different statistical input streams that are within the same sense modality. In order to provide the most optimal conditions for learning, we chose the two input streams so that they are as perceptually dissimilar as possible: colors versus shapes and tones versus nonwords.

Experiment 2: Intramodal Learning Along Different Perceptual Dimensions

The purpose of Experiment 2 is to test whether learners can learn two sets of statistical regularities when they are presented within the same sense modality but instantiated along two different perceptual "dimensions". Experiment 2A examines intramodal learning in the visual modality while Experiment 2B examines auditory learning. For Experiment 2A, one grammar is instantiated with colors and the other with shapes. For Experiment 2B, one grammar is instantiated with tones and the other with nonwords.

Method

Subjects For Experiment 2, 60 additional participants (10 in each condition) were recruited in the same manner as in Experiment 1.

Materials Experiment 2 incorporated the same two grammars, training and test sequences that were used in Experiment 1. The visual sequences were instantiated using two sets of vocabularies. The first visual vocabulary was the same set of colors as Experiment 1. The second visual vocabulary consisted of five abstract, geometric shapes. These shapes were chosen as to be perceptually distinct yet not amenable to a verbal coding strategy. The auditory sequences also were instantiated using two sets of vocabularies. The first auditory vocabulary consisted of the same set of tones as in Experiment 1. The second auditory vocabulary consisted of five different nonwords, recorded as individual sound files spoken by a human speaker (taken from Gomez, 2002): "vot", "pel", "dak", "jic", and "rud".

[1] We regard the learning as "statistical" because encoding something akin to "n-gram" chunks or transitional probabilities among sequence elements will result in above-chance test performance.

Procedure Participants were randomly assigned to one of six conditions, two for Experiment 2A, two for Experiment 2B, and two new baseline control conditions. The general procedure was the same as in Experiment 1 with the following differences. In Experiment 2A, participants were trained on both grammars, one instantiated with the color vocabulary and the other as the shape vocabulary. As in Experiment 1, participants were tested on their ability to classify novel sequences; for half of the participants, these test sequences were instantiated all as colors while for the other half they were instantiated all as shapes. Likewise, in Experiment 2B, participants were trained on both grammars, one instantiated as the tone vocabulary and the other instantiated as the nonword vocabulary. For half of these participants, the test sequences were instantiated all as tones and for the other half they were instantiated all as nonwords.

The two new baseline conditions provided data for single-grammar performance for the new shape and nonword vocabularies (note that we used the color and tone vocabulary baseline data from Experiment 1). In all other respects, the procedure for Experiment 2 was the same as in Experiment 1.

Results and Discussion

Mean scores and t-tests compared to chance levels are provided for each condition: 11.9 (59.5%), $t(9)=2.97$, $p<.05$ for the Colors-Experimental condition; 11.9 (59.5%), $t(9)=2.31$, $p<.05$ for the Shapes-Experimental condition; 13.7 (68.5%), $t(9)=4.25$, $p<.01$ for the Tones-Experimental condition; 12.0 (60.0%), $t(9)=2.58$, $p<.05$ for the Nonwords-Experimental condition; 13.2 (66.0%), $t(9)=6.25$, $p<.001$ for the Shapes-Baseline condition; and 12.2 (61.0%), $t(9)=2.34$, $p<.05$ for the Nonwords-Baseline condition. Thus, each group's overall performance was better than what would be expected by chance. Furthermore, there was no statistical difference between the respective experimental and baseline groups: Colors-Experimental versus Colors-Baseline, $t(9)=-.42$, $p=.68$; Shapes-Experimental versus Shapes-Baseline, $t(9)=-1.15$, $p=.28$; Tones-Experimental versus Tones-Baseline, $t(9)=.439$, $p=.67$; Nonwords-Experimental versus Nonwords-Baseline, $t(9)=-.178$, $p=.86$.

The results for Experiments 2A and 2B are similar to Experiment 1. Participants were adept at learning two different sets of statistical regularities simultaneously within the same sense modality, for shape and color sequences (Experiment 2A) and tone and nonword sequences (Experiment 2B). Performance levels in these dual-grammar conditions were no worse than learning levels with one grammar only. These results thus suggest that participants' learning was not mediated by abstract information. Additionally, learners can acquire statistical regularities from two streams of information within the same sense modality, at least when the two streams differ along a major perceptual dimension (colors versus shapes and tones versus nonwords). We next explore whether such learning abilities continue even when the two streams of information lie along the same perceptual dimension (two different sets of shapes and two different sets of nonwords).

Experiment 3: Intramodal Learning Within the Same Perceptual Dimension

The purpose of Experiment 3 is to test whether learners can learn two sets of statistical regularities when they are presented within the same sense modality but exist along the same perceptual "dimension". Experiment 3A incorporates two different sets of visual shapes and Experiment 3B incorporates two different sets of auditory nonwords.

Method

Subjects For Experiment 3, 60 additional participants (10 in each condition) were recruited.

Materials Experiment 3 incorporated the same two grammars, training and test sequences that were used in Experiments 1 and 2. Like the previous experiments, the experimental conditions employed learning under dual-grammar conditions. Experiment 3A employed two visual vocabularies: shape sets 1 and 2. Shape set 1 was the same set of shapes used in Experiment 2A; shape set 2 was a new set of shapes similar in overall appearance but perceptually distinct from set 1. Experiment 3B employed the nonword vocabulary used in Experiment 2B as well as a new nonword set consisting of "tood", "jeen", "gens", "tam", and "leb".

Procedure Participants were randomly assigned to one of six conditions, two for Experiment 3A, two for Experiment 3B, and two new baseline control conditions. The general procedure was identical to Experiment 2 except that different vocabularies were used. In Experiment 3A, one grammar was instantiated with shape set 1 and the other grammar was instantiated as shape set 2. At test, half of the participants were given the test sequences instantiated as shape set 1 and for the other half they were instantiated as shape set 2. Similarly, participants in Experiment 3B were also trained on both grammars, with one grammar being instantiated as nonword set 1 and the other instantiated as nonword set 2. Half of these participants were tested on the first nonword set and the other half were tested on the second nonword set.

The two new baseline conditions provided data for single-grammar performance for the new shape set 2 and nonword set 2 vocabularies (note that we used the shape set 1 and nonword set 1 baseline data from Experiment 2). In all other respects, the procedure for Experiment 3 was the same as in Experiment 2.

Results and Discussion

Mean scores and t-tests compared to chance levels are provided for each condition: 12.0 (60.0%), $t(9)=2.58$, $p<.05$ for the Shapes1-Experimental condition; 11.2 (56.0%), $t(9)=1.65$, $p=.13$ for the Shapes2-Experimental condition; 10.9 (54.5%), $t(9)=1.49$, $p=.17$ for the Nonwords1-Experimental condition; 12.4 (62.0%), $t(9)=6.47$, $p<.001$ for the Nonwords2-Experimental condition; 11.6 (58.0%), $t(9)=2.95$, $p<.05$ for the Shapes2-Baseline condition; and 13.3 (66.5%), $t(9)=3.79$, $p<.01$ for the Nonwords2-Baseline condition. We also compared each experimental group to its respective baseline performance: Shapes1-Experimental versus Shapes1-Baseline, $t(9)=-1.68$, $p=.13$; Shapes2-Experimental versus Shapes2-Baseline, $t(9)=-.89$, $p=.40$; Nonwords1-Experimental versus Nonwords1-Baseline, $t(9)=-.99$, $p=.35$; Nonwords2-Experimental versus Nonwords2-Baseline $t(9)=-.96$, $p=.36$.

Experiment 3 shows a decrement in performance of statistical learning when the two grammars are composed of vocabularies within the same perceptual dimension. When exposed to two different statistically-governed streams of visual input, each with a distinct vocabulary of shapes, learners on average are only able to learn the structure for one of the streams. This same result was also found when learners were exposed to two different nonword auditory streams. This data thus suggests that learning of multiple sources of statistical information is hindered when the input elements of the two vocabularies are perceptually similar[2]. Traditional, abstractive theories of AGL cannot account for such low-level, perceptual effects.

Overal Analyses To better quantify the differences in learning across the three experiments, we submitted all data to a 4 X 2 X 2 ANOVA that constrasted condition (crossmodal, intramodal-different dimension, intramodal-same dimension, or baseline), modality (visual versus auditory), and grammar (Grammar A versus Grammar B). There was a main effect of condition, $F(3,144)=2.66$; $p=.050$. There was a marginally significant main effect of modality, $F(1,144)=2.97$; $p=.087$. There was no main effect of grammar, $F(1,144)=1.26$; $p=.264$, nor were there any significant interactions (p's $>.05$). The marginal effect of modality is consistent with previous research showing that auditory statistical learning of sequential input is generally superior to visual or tactile learning (Conway & Christiansen, 2005).

For ease of presentation, Figure 2 shows the overall data collapsed across grammar and modality. Post-hoc comparisons reveal that the mean performance for the intramodal, same-dimension condition is significantly less than performance on both the crossmodal ($p<.01$) and baseline ($p<.05$) conditions. Thus, the ANOVA confirms that there was a learning decrement in Experiment 3, for intramodal, same-dimension learning.

General Discussion

Experiment 1 showed that learners can learn statistical regularities from two artificial grammars presented via two different input streams when they occur in different sense modalities, one visually and the other aurally. Furthermore, test performance under such dual-grammar conditions was identical to baseline, single-grammar performance. Experiments 2 and 3 extended these results, showing that learners can also learn regularities from two input streams simultaneously within the same sense modality–as long as the respective vocabularies differ along a major perceptual dimension. Learning suffered when the vocabularies for each grammar existed along the same perceptual dimension: participants could only extract statistical relationships from just one of the two input streams, not both.

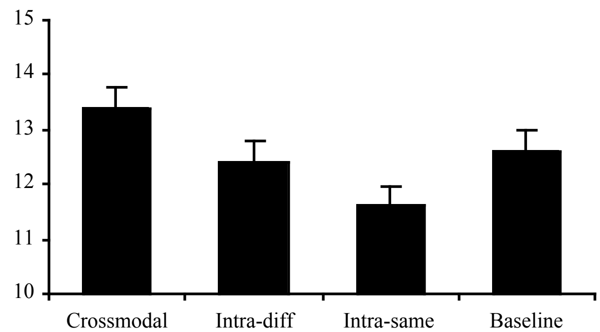

Figure 2: Mean test performance (out of 20) for all three experiments: Crossmodal (Experiment 1), Intramodal, different-dimension (Experiment 2), Intramodal, same-dimension (Experiment 3), and Baseline, single-grammar conditions (Experiments 1, 2, 3).

These studies were motivated by two questions. First, is statistical learning stimulus-specific or is it abstract and amodal? The data showed that learning was tied to the specific sense modality and perceptual dimension of the input. This stands in contrast to other arguments that learning may consist of modality-independent representations (Altmann et al., 1995) or abstract "rules" (Marcus et al., 1999; Reber, 1993)[3].

Second, can participants learn multiple, independent statistical regularities simultaneously? Quite remarkably, Experiments 1 and 2 showed that indeed they can, at least under crossmodal and intramodal (different-dimension) conditions. This ability makes sense when one considers that humans often process multiple, concurrent perceptual inputs at the same time, especially across different sensory modalities. For example, driving a car involves performing certain motor sequences as well as attending to multiple visual, auditory, and haptic input patterns. It is likely that there is an adaptive advantage for organisms to be able to encode statistical regularities from multiple environmental input streams simultaneously.

It could be that the advantage that our learners displayed for crossmodal learning may be due to attentional

[2] Another interpretation of these results is that learning in Experiment 3 was based on abstract information, leading to near-chance performance; however, it is unclear why learning would be abstract here but not so in Experiments 1 and 2.

[3] As two anonymous reviewers pointed out, an alternative possibility is that human cognition is an adaptive process relying on stimulus-specific representations in some situations and abstract learning in others.

constraints. It is known that people can better attend to rapidly-presented sequential stimuli when one stream is auditory and the other is visual, compared to when both are in the same modality (Duncan, Martens, & Ward, 1997). Thus, although it is generally assumed that implicit statistical learning does not require attention, our results indicate that attention may play an important role (also see Baker et al., 2004).

Our results also may provide insight into the underlying cognitive and neural mechanisms of statistical learning. One possibility is that statistical learning is a single mechanism that operates over all types of input (e.g., Kirkham, Slemmer, & Johnson, 2002). However, such an account has difficulty explaining the presence of learning-related modality differences (Conway & Christiansen, 2005). Furthermore, it is not clear how a single mechanism can afford simultaneous learning of multiple statistical regularities and keep the stimulus-specific representations independent of one another, as our current data show.

It may be more likely that statistical learning consists of multiple subsystems that are closely tied to specific modality-specific neural regions (Conway & Christiansen, 2005). For instance, the mismatch negativity brain response, which is elicited when a deviant sound occurs in a complex sound sequence, is generated within auditory cortex (Alho, et al., 1993). Additionally, primary and secondary visual association areas (BA 17-19) show decreased activity when participants learn complex visual patterns implicitly (Reber et al., 1998), perhaps reflecting a kind of perceptual fluency effect.

Something akin to perceptual fluency may very likely underlie statistical learning, where items that are similar in structure are processed more efficiently by networks of neurons in modality-specific brain regions (see also Chang & Knowlton, 2004). Such a view of statistical learning, and implicit learning more generally, resonates with theories of implicit memory (Schacter, Chiu, & Ochsner, 1993) and procedural learning (Goschke, 1998; Keele, Ivry, Mayr, Hazeltine, & Heuer, 2003), which also stress the involvement of multiple, modality-specific subsystems.

Conclusion

These experiments suggest that statistical learning is mediated by stimulus-specific representations. Furthermore, we've shown that learners can simultaneously encode statistical structure from two grammars originating from two different input streams and keep the knowledge representations independent of one another, as long as each is presented in a different sensory modality or along different perceptual dimensions. This suggests that the knowledge underlying statistical learning is closely tied to the perceptual features of the material itself, perhaps indicating the involvement of multiple learning subsystems.

Acknowledgments

We thank Nick Chater for helpful comments during this project.

References

Alho, K., Huotilainen, M., Tiitinen, H., Ilmoniemi, R.J., Knuutila, J., & Naatanen, R. (1993). Memory-related processing of complex sound patterns in human auditory cortex: a MEG study. *Neuroreport, 4*, 391-394.

Altmann, G.T.M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 899-912.

Baker, C.I., Olson, C.R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science, 15*, 460-466.

Chang, G.Y. & Knowlton, B.J. (2004). Visual feature learning in artificial grammar classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30*, 714-722.

Christiansen, M.H. & Curtin, S. (1999). Transfer of learning: Rule acquisition or statistical learning? *Trends in Cognitive Sciences, 3*, 289-290.

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

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

Conway, C.M. & Christiansen, M.H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences, 5*, 539-546.

Fiser, J. & Aslin, R.N. (2002). Statistical learning of higher order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 28*, 458-467.

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

Goschke, T. (1998). Implicit learning of perceptual and motor sequences: Evidence for independent learning systems. In M. Stadler & Frensch (Eds.), *Handbook of implicit learning* (pp. 401-444). Thousand Oaks, CA: Sage Publications.

Keele, S.W., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review, 110*, 316-339.

Marcus, G.F., Vijayan, S., Rao, S.B., & Vishton, P.M. (1999). Rule learning by seven-month-old infants. *Science, 283*, 77-79.

McClelland, J.L. & Plaut, D.C. (1999). Does generalization in infant learning implicate abstract algebra-like rules? *Trends in Cognitive Sciences, 3*, 166-168.

Nissen, M.J. & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*, 1-32.

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

Reber, P.J., Stark, C.E.L., & Squire, L.R. (1998). Cortical areas supporting category learning identified using functional MRI. *Proceedings of the National Academy of Sciences, USA, 95*, 747-750.

Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27-52.

Saffran, J.R., Newport, E.L., & Aslin, R.N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606-621.

Schacter, D.L., Chiu, C.Y.P., Ochsner, K.N. (1993). Implicit memory: A selective review. *Annual Review of Neuroscience, 16*, 159-182.

Seidenberg, M.S. & Elman, J.L. (1999). Networks are not 'hidden rules'. *Trends in Cognitive Sciences, 3*, 288-289.