

AN ODYSSEY THROUGH SIGHT, SOUND, AND TOUCH:
TOWARD A PERCEPTUAL THEORY OF IMPLICIT STATISTICAL
LEARNING

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Christopher M. Conway
August 2005

© 2005 Christopher M. Conway

AN ODYSSEY THROUGH SIGHT, SOUND, AND TOUCH:
TOWARD A PERCEPTUAL THEORY OF IMPLICIT STATISTICAL
LEARNING

Christopher M. Conway, Ph.D.

Cornell University, 2005

In order to steer through a world characterized by a complex mixture of variability and structure, organisms rely upon implicit statistical learning, the capability to extract probabilistic patterns occurring in environmental stimuli. Although statistical learning has been found to occur across a myriad of domains, there has been little investigation into the effect that sense modality and other stimulus attributes may have on learning. In a series of experiments, I investigate to what extent implicit statistical learning is constrained and influenced by the nature of the input in which the statistical regularities occur. All experiments have in common the use of artificial grammar learning methodology, where adult participants are incidentally exposed to statistically-governed patterns and then are tested on their ability to apply their acquired knowledge to novel instances. Chapter 2 presents two experiments that compared learning across touch, vision, and audition, producing evidence for modality constraints. Specifically, the auditory modality displayed a quantitative learning advantage compared to vision and touch; additionally, each sense modality was more or less attuned to specific aspects of the input. Chapter 3 describes an experiment that further explored modality constraints by manipulating both the presentation format (temporal, spatial, or spatiotemporal) and presentation rate for visual and auditory material. Consistent with a modality-constrained view of learning, vision and

audition were best at encoding spatial and temporal regularities, respectively. Finally, using a novel cross-over design, Chapter 4 presents three experiments that pitted abstract, amodal processing against stimulus-specific learning and found that statistical learning is mediated to a greater extent by stimulus-specific, not abstract, representations. Taken together, the results from these experiments suggest that statistical learning inherently involves learning mechanisms that are heavily influenced by the perceptual and sensory characteristics of the stimuli. I argue that a full understanding of statistical learning – and likely other aspects of language and cognition – will come only by specifying the role played by the senses. I conclude with a proposal for a perceptual, modality-constrained view of implicit statistical learning framed within the context of cognition as a whole.

BIOGRAPHICAL SKETCH

Christopher M. Conway began his doctoral graduate work at Cornell University in 2001 after receiving an MA in Psychology at Southern Illinois University for his master's thesis, *Tactile Sequential Learning: Artificial Grammar Learning by Touch*. Prior to entering graduate school, Chris completed a B.S.E. in biomedical and electrical engineering at Duke University and worked as a project engineer in a food manufacturing plant. Chris' main research interests lie in the realm of implicit statistical/sequential learning. He has presented statistical learning research at several meetings of both the Cognitive Science and Psychonomic Societies. His other research interests include animal cognition, cognitive neuroscience, and perceptual constraints on learning and memory.

ACKNOWLEDGEMENTS

I wish to thank the members of my special committee, Morten Christiansen, Michael Owren, Elizabeth Regan, and Michael Spivey, for their ongoing support and input on this research. I especially would like to thank Morten for being an outstanding mentor, advisor, and friend. His guidance was crucial to my development as a scholar and scientist and his generosity exceeded all expectations. I look forward to many more years of fruitful collaborations with him. I also wish to thank the following people for their input on various aspects of this research: Nick Chater, James Cutting, Dick Darlington, Carol Krumhansl, and Michael Young. I also wish to thank all of the Cornell Psychology faculty and students for not only creating an enriching and challenging academic environment, but also for the lasting friendships that were made, especially Rick, Emily, Kevin, Erin, Florencia, Thomas, and Luca. Finally, I would like to gratefully acknowledge the support, love, and prayers from all of my family, especially my parents, brothers John and Neil, and wife Emily.

TABLE OF CONTENTS

Biographical Sketch	iii
Acknowledgements	iv
Table of Contents	v
List of Figures	vi
List of Tables	vii
Chapter 1: Introduction	1
Chapter 2: Modality-constrained statistical learning of tactile, Visual, and auditory sequences	15
Chapter 3: Seeing and hearing in space and time: Effects of modality and presentation rate on on implicit statistical learning	76
Chapter 4: Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations	103
Chapter 5: Summary and discussion	130

LIST OF FIGURES

<i>Figure 2.1.</i> Artificial grammar used in Experiment 1.	17
<i>Figure 2.2.</i> Vibration devices used in Experiment 1A.	24
<i>Figure 2.3.</i> Graphical representation of tactile sequence.	25
<i>Figure 2.4.</i> Graphical representation of visual sequence.	34
<i>Figure 2.5.</i> Graphical representation of auditory sequence.	35
<i>Figure 2.6.</i> Experiment 1 test performance.	38
<i>Figure 2.7.</i> Artificial grammar used in Experiment 2.	46
<i>Figure 2.8.</i> Experiment 2 pretraining MDS solutions.	56
<i>Figure 2.9.</i> Experiment 2 test performance.	58
<i>Figure 3.1.</i> Artificial grammar used in Experiment 3.	80
<i>Figure 3.2.</i> Experiment 3 visual-temporal sequence.	83
<i>Figure 3.3.</i> Experiment 3 visual-spatial sequence.	84
<i>Figure 3.4.</i> Experiment 3 visual-spatiotemporal sequence.	85
<i>Figure 3.5.</i> Experiment 3 experimental group test results.	90
<i>Figure 4.1.</i> Grammars A and B used in Experiments 4-6.	107
<i>Figure 4.2.</i> Sample stimulus presentation for Experiment 4.	110
<i>Figure 4.3.</i> Visual vocabularies used in Experiment 6A.	117
<i>Figure 4.4.</i> Overall results for Experiments 4-6.	121
<i>Figure 5.1.</i> Prominent view of memory systems.	145
<i>Figure 5.2.</i> Proposed modality-constrained view of cognition.	146

LIST OF TABLES

<i>Table 2.1.</i> Experiment 1 training pairs.	26
<i>Table 2.2.</i> Experiment 1 test sequences and fragment measures.	28
<i>Table 2.3.</i> Results of Experiment 1 principal components analysis.	42
<i>Table 2.4.</i> Experiment 2 pretraining materials.	47
<i>Table 2.5.</i> Experiment 2 training materials.	48
<i>Table 2.6.</i> Experiment 2 test sequences and fragment measures.	50
<i>Table 3.1.</i> Experiment 3 training pairs.	80
<i>Table 3.2.</i> Experiment 3 test materials.	82
<i>Table 3.3.</i> Experiment 3 training results.	88
<i>Table 3.4.</i> Experiment 3 control group test results.	89
<i>Table 3.5.</i> Experiment 3 experimental group test results.	90
<i>Table 4.1.</i> Training and test items for Experiments 4-6.	108
<i>Table 4.2.</i> Experiment 4 results.	112
<i>Table 4.3.</i> Experiment 5 results.	115
<i>Table 4.4.</i> Experiment 6 results.	119

CHAPTER ONE

Introduction

Over a century ago, William James remarked that “...our capacities for forming new habits, for remembering sequences, and for abstracting general properties from things and associating their usual consequences with them...[are] exactly the faculties needed for steering us in this world of mixed variety and uniformity” (James, 1985/1892; p. xxvii). As James noted, the world is a complex mixture of variability and structure. Learning to perceive relevant environmental regularities amidst a background of change is an important function for any successfully adapting organism. For humans, detecting structure is mediated by multiple sense modalities. For instance, in the auditory domain, speech elements (e.g., words and phonemes) occur not randomly but according to certain regularities that can be described in terms of statistical or probabilistic relationships (Altmann, 2002; Christiansen, Allen, & Seidenberg, 1998; Saffran, Aslin, & Newport, 1996). That is, language, once learned, does not sound like incomprehensible gibberish; there is a lawful order to the way that speech sounds and words occur in speech. Music is another example of a structured auditory stimulus domain: music elements occur with particular temporal and rhythmic regularities (Drake & Bertrand, 2001) that are detectable by adults and infants (Hannon & Johnson, in press). Structure is also inherent in our visual environment, in the patterns of letters and words that occur in print, as well as in the way in which objects and features are arrayed in a visual scene (Fiser & Aslin, 2001; Olshausen & Field, 2000). Consider that our visual experience does not look like the chaotic “white noise” of a television screen but contains recognizable exemplars in a

non-random configuration. Finally, the tactile sense also subserves the detection of environmental regularities: tactile communication devices used by the visually impaired, such as Braille, exploit the fact that the sense of touch can usefully pick up meaningful patterns of spatiotemporal stimulus energies (Heller & Schiff, 1991).

Several decades of research have revealed that the ability to detect and use such environmental regularities is not only fundamental to human cognition and survival, but occurs relatively automatically and without conscious awareness (Barlow, 2001; Berry, 1997; Berry & Dienes, 1993; Cleeremans, Destrebecqz, & Boyer, 1998; Reber, 1993; Seger, 1994; Stadler & Frensch, 1998). This learning ability, here referred to as implicit statistical learning, has been investigated through the use of the artificial grammar learning (AGL) paradigm (Reber, 1967). In a standard AGL experiment, 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 and unsupervised learning conditions, it is the 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, though they are 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 stimuli, such as detecting element co-occurrences or covariations between events, learning can be regarded as one of computing and encoding statistically-based patterns.

The AGL paradigm has provided a rich database of knowledge regarding the characteristics of implicit statistical learning. However, one

major limitation to these studies is that the majority have used a single, narrow stimulus domain (strings of letters), making it difficult to generalize results across cognition. Researchers have mostly been content to study implicit statistical learning without regard to the effect that different stimuli may have on learning, likely due in part to two reasons. First, the field of cognition has long been enamored with theories of amodal processing, where perception is considered to be separate and independent of “higher” cognitive processes such as memory or concept formation (Fodor, 1983; Pylyshyn, 1999). Under this view, the perceptual characteristics of the stimulus elements are relatively inconsequential to the operation of the supposed amodal implicit learning mechanisms. The second reason is that, from the outset of AGL research, it was discovered that participants could apparently transfer their knowledge of the input structure from one letter set to another; that is, after being exposed to grammatical strings using one letter vocabulary, learners could then apply their knowledge of the underlying patterns to new strings formed with a different letter set (Brooks & Vokey, 1991; Gomez, 1997; Knowlton & Squire, 1996; Manza & Reber, 1997; Reber, 1969; Shanks, Johnstone, & Staggs, 1997; Whittlesea & Dorken, 1993). These transfer data suggest to some that the acquired knowledge in AGL is abstract, consisting of a high-level representation that generalizes the underlying complex input structure while necessarily excluding stimulus-specific information (Reber, 1993).

Because of these two reasons, most researchers have resisted exploring any effect that sense modality or stimulus characteristics might have on implicit statistical learning, presumably because they do not expect any effect to exist. In the handful of studies that have recently examined statistical

learning in different sensory and stimulus domains (e.g., Fiser & Aslin, 2002, 2001; Pothos & Bailey, 2000; Saffran, Johnson, Aslin, & Newport, 1999), the data have not revealed modality differences. However, such a conclusion may be premature given that these studies have not used comparable procedures and materials when comparing learning in different domains. The one notable exception is Saffran (2002), who examined AGL of visual and auditory material and did in fact report modality effects. What is needed beyond this single study is a more rigorous examination of learning across modalities and stimulus domains. A theory of implicit learning, or indeed of any aspect of cognition, will be incomplete if it does not take into account the role that sense modality plays in processing (Tulving & Madigan, 1970).

Thus, the goal of this thesis is to investigate the role that sense modality and stimulus characteristics have on implicit statistical learning. To preview, the evidence from the following experiments suggests, in contrast to most prevailing views, that the perceptual characteristics of the stimulus elements do effect learning in important ways. Chapter 2 is a recently published paper in the *Journal of Experimental Psychology: Learning, Memory, and Cognition* that explores statistical learning of tactile, visual, and auditory sequences (Conway & Christiansen, 2005), presenting the first empirical evidence of implicit statistical learning in the tactile domain. We find that although all three sense modalities have the capability for affording AGL, there exist prominent modality-specific learning differences. These learning differences took the form of what we refer to as *quantitative* and *qualitative* effects. The learning of sequential structure proceeded quantitatively better for auditory input compared to the other modalities. Additionally, each sense modality was

more or less qualitatively attuned to picking up regularities that occurred toward the beginning or ending parts of sequences.

One possibility is that this auditory learning advantage is due to more global modality constraints that affect perception, learning, and memory across the senses. I argue that audition is excellent at tasks requiring temporal processing adroitness, whereas vision excels at handling spatial configurations. Chapter 3, a paper to be submitted to the journal *Memory and Cognition*, directly explores this hypothesis by manipulating the presentation format and presentation rate in visual AGL. Visual stimuli were distributed temporally, spatially, or spatiotemporally. As predicted by our modality-constrained view, learning in the visual-spatial condition exceeded the other visual conditions, comparable to an auditory (temporal) condition. Furthermore, we found that learning of visual-temporal structure declined the most for the faster presentation rate, consistent with other research showing a similar effect in short-term memory of temporal rhythms (Collier & Logan, 2002). Finally, the qualitative biases observed earlier in Chapter 2 were replicated, supporting the notion that each learning subsystem has different properties that govern its encoding efficacy.

Even though these two papers highlight the existence of learning-related modality constraints, it may still be possible that the acquired knowledge in an AGL task consists mainly of amodal or abstract representations. To investigate this possibility, Chapter 4 presents a paper to be submitted to *Psychological Science* that pits abstract versus stimulus-specific knowledge. Using a novel modification of the standard AGL design, we presented two different sets of statistically-governed stimuli to participants, either in two different sense modalities or within the same modality. At test,

they were given new exemplars from each of the two grammars, but with all test items instantiated with one of the vocabularies only. The data showed that participants learned both sets of structured stimuli independent of one another – that is, specific to the sense modality in which it was originally presented – as long as the two sets of stimuli existed in different sense modalities (audition versus vision) or along different perceptual dimensions (nonsense syllables versus tone sequences; color sequences versus shape sequences). Thus, rather than the knowledge being in an amodal or abstract form, these experiments show that AGL is highly stimulus-specific.

Together, these three papers point toward a new account of implicit statistical learning, one that stresses sensory and perceptual constraints. In the final chapter, Chapter 5, I propose that implicit learning arises from and is continuous with perceptual processing, and as such, is heavily affected by the sensory and perceptual features of the input domain in question. I offer an outline for a perceptual theory of implicit statistical learning, guided by an embodiment perspective (e.g., Barsalou, 1999; Glenberg, 1997; Kan et al., 2003; Pecher, Zeelenberg, & Barsalou, 2004; Pulvermüller, 2001), and drawing upon data and theory from diverse research domains such as implicit memory, perceptual learning, working memory, and serial recall. In closing, I suggest a framework for understanding cognition as a whole in terms of modality-constrained systems that perform a multitude of inter-related perceptual, learning, and memorial functions.

In order to provide a foundation for the learning abilities displayed in the next chapters, I next review recent evidence investigating statistical learning in non-human primates.

Statistical Learning in Non-Human Primates

It appears that many animal species are sensitive to statistical information in the environment (Kelly & Martin, 1994). For instance, rats are highly sensitive to rate information governing the placement of food in a T-shaped maze. Over numerous trials, the rat distributes its left-right choices based on the probabilistic pattern of food placement (e.g., 70% to the left versus 30% to the right). Rats also are sensitive to the contingencies occurring between variables, demonstrated in classic conditioning experiments (Rescorla, 1988). Thus, based on the evidence from rats, it appears that non-human animals are adept at detecting statistical relationships in the environment.

However, only recently have investigators examined statistical learning in non-humans using a task comparable to that used with human adults and infants (Hauser, Newport, & Aslin, 2001; for a review of non-human sequential learning abilities more generally, see Conway & Christiansen, 2001). In the Hauser et al. study, cotton-top tamarins (*Saguinus oedipus*) first were exposed to a 20-min sequential speech stream, consisting of four different trisyllabic nonsense words (e.g., *tupiro*, *golabu*, *bidaku*, *padoti*) concatenated together in random order. The boundaries between words were not marked by any acoustic or prosodic cues. Afterwards, the tamarins were exposed to different test sound sequences and were assessed on whether they oriented toward the sound when it was played. Some of the test sequences were words that were contained within the speech stream, some were non-words which contained syllables in an order that had not occurred in the speech stream, and others were part-words, which contained syllable sequences spanning a word boundary. The tamarins were significantly more

likely to orient toward non-words than to words, suggesting that they had discriminated test sequences on the basis of syllable order. The tamarins also were significantly more likely to orient toward part-words than to words, indicating that they were sensitive to the frequency of the syllable combinations. These results, which mirrored those of human infants (Saffran et al., 1996), indicate that cotton-top tamarins--and presumably other primate species--are able to encode some of the statistical regularities present in language-like, auditory sequences. However, it is important to note that while the tamarins were exposed to a 20-min speech stream, the human infants demonstrated statistical learning after only a 2-min exposure.

Two other studies have further illuminated primate statistical learning abilities. Newport, Hauser, Spaepen, and Aslin (2004) showed that tamarin monkeys can learn statistical regularities between non-adjacent elements. This is an intriguing demonstration because human language consists of regularities that occur not just among adjacent speech elements but also between elements that are separated by intervening material. A second study showed that tamarin monkeys can generalize patterns that they have learned to a new vocabulary (Hauser, Weiss, & Marcus, 2002). It has been argued that extracting these sorts of patterns is necessary for language acquisition (Marcus, Vijayan, Rao, & Vishton, 1999).

In summary, these studies appear to support the notion that statistical learning is not limited to human cognition. Non-human primates and likely other non-human animals appear to have many of the same statistical learning capabilities as humans. However, it is likely that species-specific differences exist, as well. For example, Newport et al. (2004) found differences in the types of non-adjacent dependencies that were learned by human adults and tamarin

monkeys. Thus, future research will be necessary to explore just how species-general this capability is.

References

- Altmann, G.T. (2002). Statistical learning in infants. *Proceedings of the National Academy of Sciences*, 99, 15250-15251.
- Barlow, H. (2001). The exploitation of regularities in the environment by the brain. *Behavioral and Brain Sciences*, 24, 602-607.
- Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-660.
- Berry, D.C. (1997). *How implicit is implicit learning?* Oxford: Oxford University Press.
- Berry, D.C., & Dienes, Z. (1993a). *Implicit learning: Theoretical and empirical issues*. Hove (UK): Lawrence Erlbaum, Associates.
- Brooks, L.R., & Vokey, J.R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General*, 120, 316-323.
- Christiansen, M.H., Allen, J., & Seidenberg, M.S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, 13, 221-268.
- Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, 2, 406-416.
- 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.

- Drake, C. & Bertrand, D. (2001). The quest for universals in temporal processing in music. In R.J. Zatorre & I. Peretz (Eds), *The biological foundations of music* (pp. 17-27). New York: New York Academy of Sciences.
- 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.
- Fiser, J. & Aslin, R.N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12, 499-504.
- Fodor, J.A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.
- Glenberg, A.M. (1997). *What memory is for*. Behavioral and Brain Sciences, 20, 1-55.
- Gomez, R.L. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology*, 33, 154-207.
- Hannon, E.E. & Johnson, S.P. (in press). Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. *Cognitive Psychology*.
- Hauser, M.D., Weiss, D., & Marcus, G. (2002). Rule learning by cotton-top tamarins. *Cognition*, 86, B15-B22.
- Heller, M.A., & Schiff, W. (1991a). *The psychology of touch*. Hillsdale, NJ: Lawrence Erlbaum, Associates.
- Hauser, M.D., Newport, E., & Aslin, R.N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, 78, B53-B64.
- James, W. (1985/1892). *Psychology: The briefer course*. University of Notre Dame Press.

- Kan, I.P., Barsalou, L.W., Solomon, K.O., Minor, J.K., & Thompson-Schill, S.L. (2003). Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. *Cognitive Neuropsychology*, 20, 525-540.
- Kelly, M.H., & Martin, S. (1994). Domain-general abilities applied to domain-specific tasks: Sensitivity to probabilities in perception, cognition, and language. *Lingua*, 92, 105-140.
- Manza, L., & Reber, A.S. (1997). Representing artificial grammars: Transfer across stimulus forms and modalities. In D.C. Berry (Ed.), *How implicit is implicit learning?* (pp. 73-106). Oxford University.
- Marcus, G.F., Vijayan, S., Rao, S.B., & Vishton, P.M. (1999). Rule learning by seven-month-old infants. *Science*, 283, 77-79.
- Newport, E.L., Hauser, M.D., Spaepen, G., & Aslin, R.N. (2004). Learning at a distance II: Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, 49, 85-117.
- Olshausen, B.A. & Field, D.J. (2000). Vision and the coding of natural images. *American Scientist*, 88, 238-245.
- Pecher, D., Zeelenberg, R., & Barsalou, L.W. (2004). Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation. *Psychonomic Bulletin & Review*, 11, 164-167.
- Pothos, E.M., & Bailey, T.M. (2000). The role of similarity in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 847-862.
- Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, 5, 517-524.

- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case of cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22, 341-423.
- Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855-863.
- Reber, A.S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, 81, 115-119.
- Reber, A.S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford: Oxford University Press.
- Rescorla, R.A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, 43, 151-160.
- Saffran, J.R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, 47, 172-196.
- Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27-52.
- Seger, C.A. (1994). Implicit learning. *Psychological Bulletin*, 115, 163-196.
- Shanks, D.R., Johnstone, T., & Staggs, L. (1997). Abstraction processes in artificial grammar learning. *Quarterly Journal of Experimental Psychology*, 50, 216-252.
- Stadler, M.A. & Frensch, P.A. (Eds.) (1998). *Handbook of implicit learning*. London: Sage Publications.
- Tulving, E. & Madigan, S.A. (1970). Memory and verbal learning. *Annual Review of Psychology*, 21, 437-484.

Whittlesea, B.W., & Dorken, M.D. (1993). Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General*, 122, 227-248.

CHAPTER TWO

Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences*

The world is temporally bounded: Events do not occur all at once but rather are distributed in time. Therefore, it is crucial for organisms to be able to encode and represent temporal order information. One potential method for encoding temporal order is to learn the statistical relationships of elements within sequential input. This process appears to be important in a diverse set of learning situations, including speech segmentation (Saffran, Newport, & Aslin, 1996), learning orthographic regularities of written words (Pacton, Perruchet, Fayol, & Cleeremans, 2001), visual processing (Fiser & Aslin, 2002), visuomotor learning (e.g., serial reaction time tasks; Cleeremans, 1993) and nonlinguistic auditory processing (Saffran, Johnson, Aslin, & Newport, 1999). Not only human adults but also infants (Gomez & Gerken, 1999; Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996) and nonhuman primates (Hauser, Newport, & Aslin, 2001) are capable of statistical learning.

Noting such widespread examples of statistical learning, many researchers—either implicitly or explicitly—view statistical learning as a single, domain-general phenomenon (e.g., Kirkham et al., 2002). Although it may be true that statistical learning across different domains is based on similar computational principles, it is also likely that modality constraints exist that may differentially affect such processing. For instance, traditionally,

* 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. Copyright © 2005 by the American Psychological Association. Adapted with permission.

vision and audition have been viewed as spatial and temporal senses, respectively (Kubovy, 1988). Empirical evidence from perceptual and temporal processing experiments supports such a distinction between vision and audition (e.g., Glenberg & Swanson, 1986; Mahar, Mackenzie, & McNicol, 1994). However, it is currently unknown whether and how these modality constraints affect the learning of statistical relationships between elements contained within sequential input.

This article explores potential modality constraints affecting statistical learning. Experiment 1 investigates statistical learning in three sensory modalities: touch, vision, and audition. Experiment 1A provides the first direct evidence that touch can mediate statistical learning. Experiments 1B and 1C compare learning in two additional sensory modalities, vision and audition. Although commonalities exist, we find initial evidence for a striking difference in auditory statistical learning compared with tactile and visual learning. We follow up with Experiment 2, designed to control perceptual and training effects as well as to tease apart potential learning sensitivities uncovered in the first experiment. The results of Experiment 2 provide further evidence that modality constraints affect statistical learning. We discuss these results in relation to basic issues of cognitive and neural organization—namely, to what extent statistical learning might consist of a single or multiple neural mechanisms.

Statistical Learning of Sequential Input

Statistical learning appears to be a crucial learning ability. For instance, making sense of visual scenes may require the extraction of statistical components (e.g., Fiser & Aslin, 2001). Another domain in which statistical

learning likely plays an important role is the encoding of sequential input (Conway & Christiansen, 2001). Artificial grammar learning (AGL; Reber, 1967) is a paradigm widely used for studying such statistical learning¹. AGL experiments typically use finite-state grammars to generate the stimuli. In such grammars, a transition from one state to the next produces an element of the sequence. For example, by passing through the nodes S1, S2, S2, S4, S3, S5 of Figure 2.1, one generates the “legal” sequence 4–1–3–5–2.

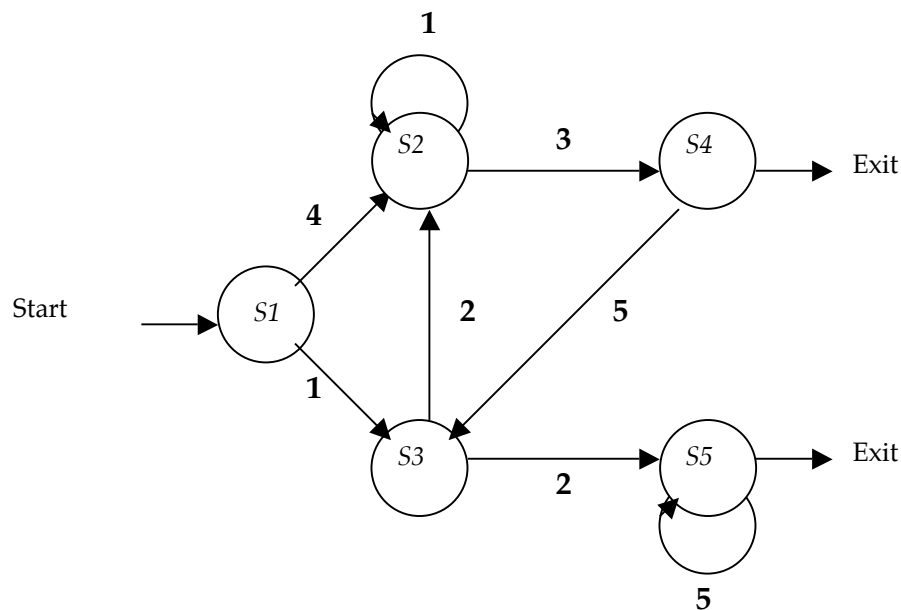


Figure 2.1. Artificial grammar from Gomez and Gerken (1999), also used in the current Experiment 1. We generated legal sequences by following the paths starting at S1 and continuing until we reached an exit path. Each path generates a number (1, 2, 3, 4, or 5) that corresponds to a particular stimulus element. S = state, so that S1 and S2 refer to State 1 and State 2, and so on.

In the AGL paradigm, participants observe a subset of legal training sequences (i.e., sequences that are generated from the artificial grammar), after which the participants typically display learning of sequential structure as

¹ The serial reaction time (SRT) task is another common method for exploring the learning of sequential regularities. The SRT paradigm differs from AGL in that the behavioral measure for the former is reaction time, whereas that for the latter is classification accuracy.

evidenced by their ability to classify novel sequences as being legal or illegal. Additionally, they often have difficulties verbalizing the distinction between legal and illegal stimuli, a finding that originally prompted Reber (1967) to describe the learning as implicit.

The nature of the cognitive processes underlying AGL has been the subject of much debate, leading to the proposal of several different theories. The abstractive view sees AGL as a process that encodes and extracts the abstract rules of the grammar (e.g., Reber, 1993). Two alternative accounts stand in contrast to the abstractive view, proposing that instead of abstract knowledge, participants learn particular features of the training items. The exemplar-based view posits that the stimuli themselves are encoded and stored in memory (e.g., Vokey & Brooks, 1992): When participants make classification judgments at test, they compare the test sequences with their memory of the stored exemplars and make their decision on the basis of similarity. The fragment-based view posits that participants learn small fragments or chunks of information, consisting of pairs (*bigrams*) and triples (*trigrams*) of elements (e.g., Perruchet & Pacteau, 1990). Participants use these chunks of information to help them classify novel input.

Although there has been disagreement as to which theory is correct, there is considerable evidence suggesting that the learning of fragment information is a crucial aspect of AGL² (e.g., Johnstone & Shanks, 1999; Knowlton & Squire, 1994, 1996; Meulemans & Van der Linden, 1997; Perruchet & Pacteau, 1990; Pothos & Bailey, 2000; Redington & Chater, 1996). These experiments have shown that participants become sensitive to the fragment

² It also appears to be the case that learners rely on other cues, such as overall similarity of test items to training exemplars, in addition to fragment information (e.g., see Pothos & Bailey, 2000).

information contained within the training input, as quantified by specific fragment measures, which allows participants to classify novel sequences in terms of whether they conform to the same statistical regularities as the training items. Such statistical sensitivity appears to be vital for AGL tasks.

The standard AGL paradigm has been used extensively to assess visual as well as auditory (e.g., Saffran, 2000) learning. However, two issues remain relatively unexplored: Can statistical learning occur in other modalities, such as touch? And what differences in statistical learning, if any, exist among different sensory modalities? Whereas previous research generally has focused on the similarities among statistical learning in different domains (Fiser & Aslin, 2002; Kirkham et al., 2002), there are reasons to suppose that modality constraints may affect learning across the various senses. Next, we summarize evidence for such modality constraints.

Modality Constraints

Ample research testifies to the existence of modality constraints that affect the manner in which people perceive, learn, and represent information (for relevant reviews, see Freides, 1974; Penney, 1989). In this section we summarize research in the realms of serial recall, temporal acuity, and the learning of temporal and statistical patterns.

One of the most well-known modality effects—often referred to as *the* modality effect—is found in serial recall. Numerous studies attest to differences in the serial position learning curves for aurally versus visually presented verbal input (e.g., lists of spoken or written words). Specifically, there appears to be a stronger recency effect (i.e., better recall of final elements in a list) for auditory as compared with visual material (Crowder, 1986; Engle

& Mobley, 1976). A number of theories have attempted to explain this modality effect, such as the traditional account supposing that a precategorical acoustic storage exists for auditory material (Crowder & Morton, 1969) or that the auditory modality benefits from better temporal coding (e.g., Glenberg & Fernandez, 1988). Beaman (2002) showed that under certain conditions, a stronger primacy effect (i.e., better recall of beginning elements in a list) occurs for visual as compared with auditory material. Traditional theories do not adequately explain why this might occur. Additionally, studies with nonhuman primates have shown that monkeys have opposite serial position curves for auditory and visual material (Wright, 2002), as a function of the amount of time occurring between the last element in the list and the recall test. That is, when the recall test occurs relatively soon after the list presentation, there is an auditory primacy effect and a visual recency effect; when the recall test occurs relatively late after the presentation, there is a visual primacy and an auditory recency effect. These new data suggest that different mechanisms may underlie auditory and visual serial recall, leading to qualitatively different serial position curves.

Modality differences are also apparent in low-level temporal processing tasks (e.g., Gescheider, 1966, 1967; Lechelt, 1975; Oatley, Robertson, & Scanlan, 1969; Sherrick & Cholewiak, 1986). For example, Sherrick and Cholewiak (1986) reviewed data relating to temporal acuity in touch, vision, and audition. In measures of simultaneity—the ability to correctly perceive two closely occurring events—the senses have differing temporal sensitivity, with vision being the least and audition the most sensitive. Similarly, Lechelt (1975) assessed each modality in terms of numerosity, or the ability to count rapidly presented stimuli. Stimuli consisting of flashes of light, aural clicks, or finger

taps were delivered for short durations (2 ms or less), with sequences of varying length (between two and nine pulses) and varying rates (between three and eight signals per second). In terms of assessing the number of signals in the sequences, participants performed best when the signals were presented aurally and worst when they were presented visually.

Likewise, studies of temporal pattern and rhythm discrimination also reveal modality differences (e.g., Collier & Logan, 2000; Garner & Gottwald, 1968; Glenberg & Jona, 1991; Handel & Buffardi, 1969; Manning, Pasquali, & Smith, 1975; Rubinstein & Gruenberg, 1971). When presented with rhythmic patterns of flashing lights or auditory stimuli, participants were much better at discriminating auditory as opposed to visual patterns (Rubinstein & Gruenberg, 1971). Learners were also better at identifying repeating sequences of binary elements (e.g., 1122121211221212) when the elements were auditory stimuli rather than visual or tactual ones (Handel & Buffardi, 1969).

There have also been hints that similar modality constraints affect AGL. Several studies have noted that performance in AGL tasks differs depending on the modality and the manner of presentation (i.e., whether material is presented simultaneously or sequentially). For instance, Gomez (1997) remarked that visual AGL proceeds better when the stimuli are presented simultaneously rather than sequentially, perhaps because a simultaneous format permits better chunking of the stimulus elements. Saffran (2002) used an AGL task to test participants' ability to learn predictive dependencies. She found that participants learned these predictive relationships best with an auditory-sequential or visual-simultaneous presentation and did poorly in a visual-sequential condition.

The evidence reviewed suggests that modality differences are present across the cognitive spectrum. These modality constraints take two main forms. First, it appears that vision and audition differ in respect to their sensitivities to the initial or final parts of sequential input. Vision may be more sensitive to initial items in a list (Beaman, 2002), whereas audition appears more sensitive to final list items (Crowder, 1986). Second, the auditory modality appears to have an advantage in the processing of sequential input, including low-level temporal processing tasks (Sherrick & Cholewiak, 1986) and pattern or rhythm discrimination (e.g., Manning et al., 1975). In a comprehensive review of the effect of modality on cognitive processing, Friedes (1974) concluded that for complex tasks, audition is best suited for temporal processing, whereas vision excels at spatial tasks (for similar views, see also Kubovy, 1988; Mahar et al., 1994; Penney, 1989; Saffran, 2002). That is, audition is best at processing sequential, temporally distributed input, whereas vision excels at spatially distributed input. The touch modality appears to be adept at processing both sequential and spatial input, but not at the same level of proficiency as either audition or vision (Mahar et al., 1994).

In this article we explore in what manner these modality constraints might affect statistical learning. In the experiments, our strategy is to incorporate comparable input in three sensory conditions: touch, vision, and audition. Previous researchers have claimed that statistical learning in audition and vision is the same, yet rarely has much effort been made to control experimental procedures and materials across the senses. Thus, the present experiments provide a better comparison of learning across these three modalities. We begin by investigating statistical learning in the tactile domain, a realm that has been previously ignored in AGL experiments.

Experiment 1A: Tactile Statistical Learning

The touch sense has been studied extensively in terms of its perceptual and psychophysical attributes (see Craig & Rollman, 1999), yet it has not been fully explored in relation to statistical learning. In Experiment 1A, we presented to participants tactile sequences conforming to an artificial grammar and then tested their ability to classify novel sequences. As reviewed above, studies of sequential pattern perception suggest that the touch sense ought to be capable of extracting sequential regularities in an AGL setting (e.g., Handel & Buffardi, 1969; Manning et al., 1975). This experiment attempted to verify this hypothesis.

Method

Participants

Twenty undergraduates (10 in each condition) from introductory psychology classes at Southern Illinois University participated in the experiment. Subjects earned course credit for their participation. The data from an additional 5 participants were excluded for the following reasons: prior participation in AGL tasks in our laboratory ($n = 4$), and failure to adequately follow the instructions ($n = 1$).

Apparatus

The experiment was conducted with the PsyScope presentation software (Cohen, MacWhinney, Flatt, & Provost, 1993) run on an Apple G3 PowerPC computer. Participants made their responses using an input/output button box (New Micros, Inc., Dallas, TX). Five small motors (18 mm × 5 mm), normally used in hand-held paging devices, generated the vibrotactile pulses (rated at 150 Hz). The vibration pulses were suprathreshold stimuli and easily

perceived by all participants. The motors were controlled by output signals originating from the New Micros button box. These control signals were in turn determined by the PsyScope program, which allowed precise control over the timing and duration of each vibration stimulus. Figure 2.2 shows the general experimental setup.

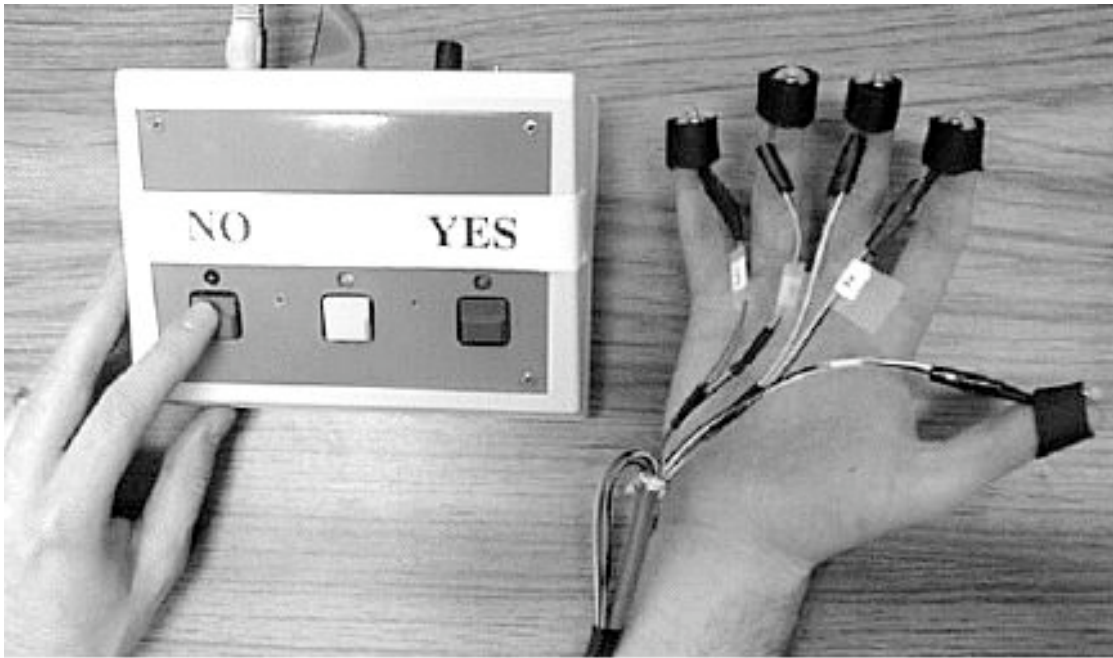


Figure 2.2. Vibration devices attached to a participant's hand with the button box to the side (Experiment 1A)

Materials

The stimuli used for Experiment 1 were taken from Gomez and Gerken's (1999) Experiment 2. This grammar (see Figure 2.1) can generate up to 23 sequences between three and six elements in length. The grammar generates sequences of numbers. Each number from the grammar was mapped onto a particular finger (1 was the thumb, and 5 was the pinky finger). Each sequence generated from the grammar thus represents a series of

vibration pulses delivered to the fingers, one finger at a time. Each finger pulse duration was 250 ms, and the pulses within a sequence were separated by 250 ms. As an illustration, the sequence 1–2–5–5 corresponds to a 250-ms pulse delivered to the thumb, a 250-ms pause, a 250-ms pulse delivered to the second finger, a 250-ms pause, a 250-ms pulse delivered to the fifth finger, a 250-ms pause, and then a final 250-ms pulse delivered to the fifth finger.

Figure 2.3 graphically represents this sequence.

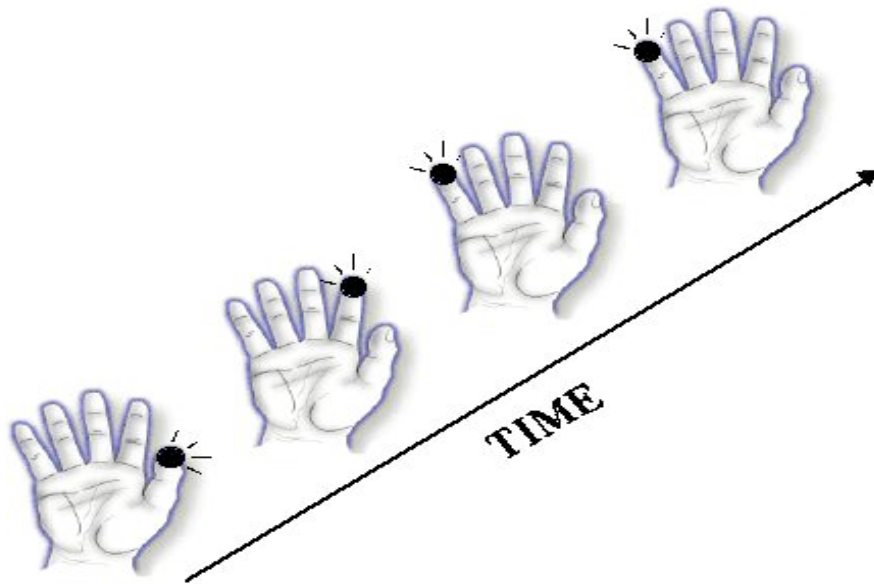


Figure 2.3. Graphical representation of the tactile sequence 1–2–5–5 in Experiment 1A. Each hand represents a single slice in time, whereas each black circle represents the occurrence of a vibrotactile pulse to a particular finger.

A total of 12 legal sequences were used for training.³ Each of the legal sequences was used twice to formulate a set of 12 training pairs. Six pairs consisted of the same training sequence presented twice (matched pairs),

³ Note that what we refer to as the training phase contained neither performance feedback nor reinforcement of any kind. Exposure phase might be a more accurate description of this part of the experiment.

whereas the remaining 6 pairs consisted of 2 sequences that differed slightly from one another (mismatched pairs). These matched and mismatched training pairs were used in conjunction with a same–different judgment task, described in detail below. The 12 training pairs are listed in Table 2.1.

Table 2.1

Training Pairs for Experiments 1A, 1B, and 1C

Matched pairs	Mismatched pairs
1-2-1-1-1-3 / 1-2-1-1-1-3	1-2-3-5-2-5 / 1-2-3-5-2-3
4-1-1-3-5-2 / 4-1-1-3-5-2	1-2-3-5-2-3 / 1-2-3-5-2-5
4-1-3-5-2 / 4-1-3-5-2	4-3-5-2-3 / 4-3-5-2-5
1-2-5-5-5 / 1-2-5-5-5	4-3-5-2-5 / 4-3-5-2-3
4-1-3 / 4-1-3	1-2-5-5 / 1-2-1-3
1-2-3 / 1-2-3	1-2-1-3 / 1-2-5-5

Note. The numbers refer to a particular finger vibration (Experiment 1A), visual stimulus (Experiment 1B), or auditory tone (Experiment 1C.).

The test set consisted of 10 novel legal and 10 illegal sequences. Legal sequences were produced from the finite-state grammar in the normal fashion. Illegal sequences did not conform to the regularities of the grammar. The illegal sequences each began with a legal element (i.e., 1 or 4), followed by one or more illegal transitions and ending with a legal element (i.e., 2, 3, or 5). For example, the illegal sequence 4–2–1–5–3 begins and ends with legal elements (4 and 3, respectively) but contains several illegal interior transitions (4–2, 1–5, and 5–3, combinations of elements that the grammar does not allow). Therefore, the legal and illegal sequences can be described as differing from

one another in terms of the statistical relationships between adjacent elements. That is, a statistical learning mechanism able to encode the possible element combinations occurring in the training set could discern which novel test sequences are illegal. For instance, by realizing that the elements 4 and 2 never occur together in the training set, a learner could potentially discern that the novel test sequence 4–2–1–5–3 is illegal.⁴ Finally, the legal and illegal test sequences were closely matched in terms of element frequencies and sequence lengths (Gomez & Gerken, 1999). All test sequences are listed in Table 2.2.

Procedure

Participants were assigned randomly to either a control group or an experimental group. The experimental group participated in both a training and a test phase, whereas the control group only participated in the test phase. Before beginning the experiment, all participants were assessed by the Edinburgh Handedness Inventory (Oldfield, 1971) to determine their preferred hand. The experimenter then placed a vibration device onto each of the five fingers of the participant's preferred hand. At the beginning of the training phase, the experimental group participants were instructed that they were participating in a sensory experiment in which they would feel pairs of vibration sequences. For each pair of sequences, they had to decide whether the two sequences were the same and indicate their decision by pressing a button marked *YES* or *NO*. This match–mismatch paradigm used the 12 training pairs described earlier, listed in Table 2.1. It was our intention that this paradigm would encourage participants to pay attention to the stimuli

⁴ Note that we remain neutral as to whether such performance might occur in the presence or absence of awareness.

while not directly tipping them off to the nature of the statistically governed sequences.

Table 2.2

Fragment Measures for Experiments 1A, 1B, and 1C Test Sequences

Item	Chunk	Legal Test Sequences				
		Novel	NFP	Sim	I-Anchor	F-Anchor
4-1-3-5-2-3	4.11	0	0	2	2.5	2.5
1-2-1-3-5-2	4.11	0	0	2	4.5	2.0
4-3-5-2-5-5	3.67	0	2	4	2.0	1.5
4-1-3-5-2-5	4.00	0	0	2	2.5	2.0
4-1-1-1-3	2.57	0	3	3	2.0	2.0
1-2-1-1-3	3.14	0	2	2	4.5	2.0
1-2-3-5-2	5.0	0	0	1	5.0	2.0
4-1-1-3	2.80	0	0	2	2.0	2.0
4-3-5-2	4.40	0	0	1	2.0	2.0
1-2-5	4.33	0	0	1	4.5	1.0
Averages:	3.81	0.0	0.70	2.00	3.15	1.9

Table 2.2 (Continued)

		Illegal Test Sequences				
1-4-5-1-3-3	0.56	8	9	3	0	0
4-5-1-2-1-3	1.89	4	8	3	0	2.0
4-2-1-3-1-5	0.89	6	6	3	0	0
1-5-3-3-2-2	0.00	9	9	3	0	0
1-5-3-4-2	0.00	7	7	3	0	0
4-2-1-5-3	0.29	6	6	3	0	0
1-5-3-1-2	1.00	6	7	3	0	0
4-5-1-3	1.00	4	4	2	0	1.5
4-5-2-2	1.20	4	5	3	0	0
1-4-2	0.00	3	3	2	0	0
Averages:	0.68	5.70	6.40	2.80	0	0.35

Note. NFP = Novel fragment position; Sim = similarity; I-anchor = initial anchor strength; F-anchor = final anc. str.

Each pair was presented six times in random order for a total of 72 exposures. As mentioned earlier, all vibration pulses had a duration of 250 ms and were separated by 250 ms within a sequence. A 2-s pause occurred between the two sequences of each pair and after the last sequence of the pair. A prompt was displayed on the computer monitor asking for the participant's response, and it stayed on the screen until a button press was made. After another 2-s pause, the next training pair was presented. The entire training phase lasted roughly 10 min for each participant.

A recording of white noise was played during training to mask the sounds of the vibrators. In addition, the participants' hands were occluded so

that they could not visually observe their fingers. These precautions were taken to ensure that tactile information alone, without help from auditory or visual senses, contributed to task performance.

Before the beginning of the test phase, the experimental group participants were told that the vibration sequences they had just felt had been generated by a computer program that determined the order of the pulses by using a complex set of rules. They were told that they would now be presented with new vibration sequences. Some of these would be generated by the same program, whereas others would not be. It was the participant's task to classify each new sequence accordingly (i.e., whether or not the sequence was generated by the same rules) by pressing a button marked either *YES* or *NO*. The control participants, who did not participate in the training phase, received an identical test task.

The 20 test sequences were presented one at a time, in random order, to each participant. The timing of the test sequences was the same as that used during the training phase (250-ms pulse duration, 250-ms interstimulus interval, and 2-s pauses before and after each sequence). The white noise recording and occluding procedures also were continued in the test phase.

At the completion of the experiment, participants were asked how they decided whether test sequences were legal or illegal. Some researchers have used such verbal reports as a preliminary indication as to whether learning proceeded implicitly or explicitly (Seger, 1994).

Results and Discussion

We assessed the training performance for the experimental participants by calculating the mean percentage of correctly classified pairs. Participants, on average, made correct match–mismatch decisions for 74% of the training trials.

However, for our purposes, the test results are of greater interest because here the participants must generalize from training experience to previously unobserved test sequences. The control group correctly classified 45% of the test sequences, whereas the experimental group correctly classified 62% of the test sequences. Following Redington and Chater's (1996) suggestions, we conducted two analyses on the test data. The first was a one-way analysis of variance (ANOVA; experimental vs. control group) to determine whether any differences existed between the two groups. The second compared performances for each group with hypothetical chance performance (50%) using single group *t* tests.

The ANOVA revealed that the main effect of group was significant, $F(1, 18) = 3.16, p < .01$, indicating that the experimental group performed significantly better than the control group. Single group *t* tests confirmed the ANOVA's finding. The control group's performance was not significantly different from chance, $t(9) = 1.43, p = .186$, whereas the experimental group's performance was significantly above chance, $t(9) = 2.97, p < .05$.

Finally, the participants' verbal reports suggest that they had very little explicit knowledge concerning sequence legality. Most of the experimental group participants reported basing their responses merely on whether a sequence felt familiar or similar. Several of the participants reported that they

made their judgments on the basis of a simple rule (e.g., “If a sequence was four elements long, I said ‘no’”). However, in each of these cases, following the rule would actually lead to incorrect judgments. None of the participants was able to report anything specific that could actually help him or her make a decision (e.g., “Certain finger combinations were not allowed, such as the fourth finger followed by the second”). On the basis of these verbal reports, we do not see evidence that the experimental group participants were explicitly aware of the distinction between legal and illegal sequences.⁵

The results show that the experimental group significantly outperformed the control group. This suggests that the experimental participants learned aspects of the statistical structure of the training sequences—in the form of adjacent element co-occurrence statistics—that allowed them to classify novel test sequences appropriately. Additionally, the participants had difficulty verbalizing the nature of sequence legality. This is the first empirical evidence of an apparently implicit, tactile statistical learning capability.

Experiments 1B and 1C: Visual and Auditory Statistical Learning

Experiment 1A showed that statistical learning can occur in the tactile domain. To compare tactile with visual and auditory learning, we conducted two additional studies. Experiments 1B and 1C assessed statistical learning in the visual and auditory domains, respectively, using the same general procedure and statistically governed stimulus set as used in Experiment 1A. For Experiment 1B, the sequences consisted of visual stimuli occurring at

⁵ We note, however, that verbal reports are not necessarily the most sensitive measure of explicit awareness, so it is still possible that explicit awareness contributed to task performance.

different spatial locations. For Experiment 1C, sequences of tones were used. Like the vibrotactile sequences, the visual and auditory stimuli were nonlinguistic, and thus participants could not rely on a verbal encoding strategy.

Method

Participants

Experiment 1B. Twenty undergraduates (10 in each condition) were recruited from introductory psychology classes at Cornell University. Subjects received extra credit for their participation. The data from 3 additional participants were excluded because the participants did not adequately follow the instructions ($n = 2$) and because of equipment malfunction ($n = 1$).

Experiment 1C. An additional 20 undergraduates (10 in each condition) were recruited from introductory psychology classes at Cornell University.

Apparatus

The apparatus was the same as in Experiment 1A, except for the exclusion of the vibration devices. The auditory stimuli were generated by the SoundEdit 16 (Version 2) software for the Macintosh.

Materials

The training and test materials were identical to those of Experiment 1A (see Tables 2.1 and 2.2). The difference was that the sequence elements were mapped onto visual or auditory stimuli instead of vibrotactile pulses. For Experiment 1B, the stimuli consisted of black squares displayed on the computer monitor in different locations (the element 1 represents the leftmost location, and 5 the rightmost). Each black square (2.6×2.6 cm) was positioned in a horizontal row across the middle of the screen at approximately eye level,

with 2.5 cm separating each position. Participants were seated at a viewing distance of approximately 45 cm to 60 cm from the monitor.

A visual stimulus thus consisted of a spatiotemporal sequence of black squares appearing at various locations. As in Experiment 1A, each element appeared for 250 ms, and each was separated by 250 ms. Figure 2.4 shows a representation of the sequence 1-2-5-5.

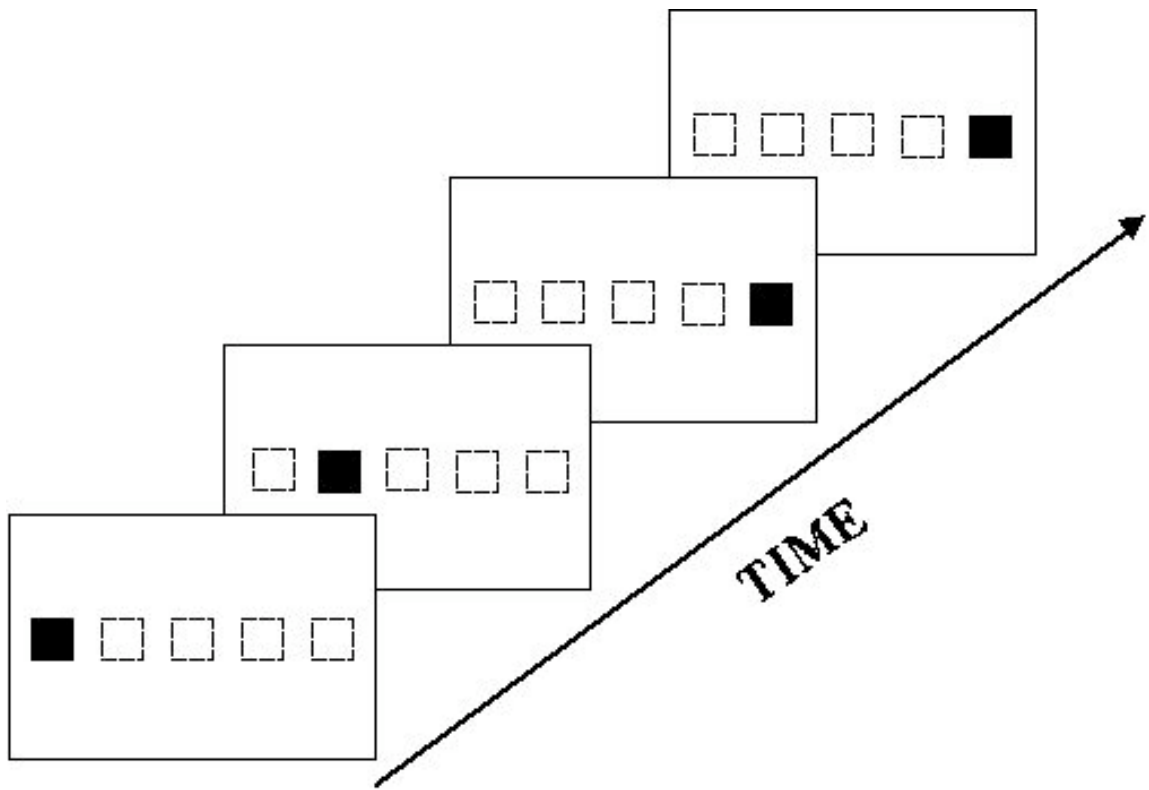


Figure 2.4. Graphical representation of the visual sequence 1-2-5-5 in Experiment 1B. Each of the four large rectangles represents the monitor display at a single slice in time. Note that the dashed squares, representing the five possible stimulus element locations, were not visible to the participants.

For Experiment 1C, the stimuli consisted of pure tones of various frequencies (1 = 261.6 Hz, 2 = 277.2 Hz, 3 = 349.2 Hz, 4 = 370 Hz, and 5 = 493.9 Hz) corresponding to musical notes C, C#, F, F#, and B, respectively.⁶ As in Experiments 1A and 1B, each element (tone) lasted 250 ms, and each was separated by 250 ms. Figure 2.5 graphically represents the sequence 1–2–5–5.

Procedure

The procedures were the same as that of Experiment 1A, the only differences relating to the nature of the stimulus elements, as described above. The timing of the stimuli, pauses, and prompts was identical to the timing in Experiment 1A.

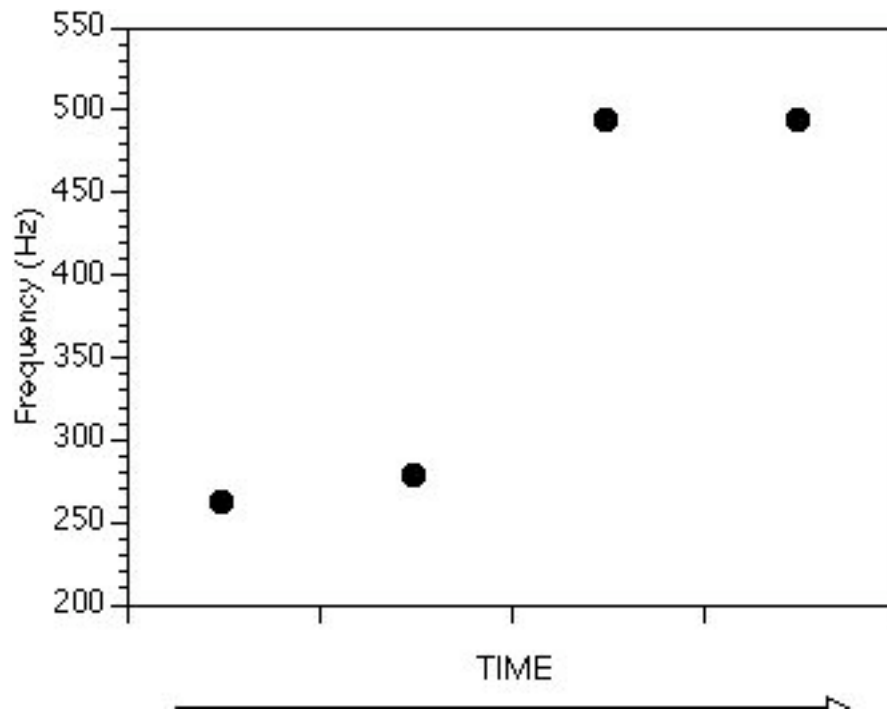


Figure 2.5. Graphical representation of the auditory sequence 1–2–5–5 in Experiment 1C.

⁶ This particular set of tones was used because it avoids familiar melodies (Dowling, 1991).

Results

We performed the same statistical analyses as used in Experiment 1A. During the training phase, the Experiment 1B (visual) experimental group made correct match–mismatch decisions on 86% of the trials, whereas the Experiment 1C (auditory) experimental group scored 96%. We compared the training means across the three experiments, which revealed a main effect of modality, $F(2, 27) = 24.30, p < .0001$. Thus, auditory training performance was significantly better than visual performance ($p < .005$), which in turn was significantly better than tactile performance ($p < .001$). Because the training task essentially involves remembering and comparing sequences within pairs, the results may elucidate possible differences among the three modalities in representing and maintaining sequential information (Penney, 1989). It is also possible that these results instead are due to factors such as differential discriminability or perceptibility of sequence elements in different sensory domains.

Results for the test phase in Experiment 1B revealed that the control group correctly classified 47% of the test sequences, whereas the experimental group correctly classified 63% of the test sequences. An ANOVA (experimental vs. control group) indicated that the main effect of group was significant, $F(1, 18) = 3.15, p < .01$. Single group t tests revealed that the control group's performance was not significantly different from chance, $t(9) = 1.11, p = .3$, whereas the experimental group's performance was significantly different from chance, $t(9) = 3.03, p < .05$.

Results for the auditory (Experiment 1C) test phase revealed that the control group correctly classified 44% of the test sequences, whereas the experimental group correctly classified 75% of the test sequences. An ANOVA

(experimental vs. control group) indicated that the main effect of group was significant, $F(1, 18) = 7.08, p < .001$. Single group t tests revealed that the control group's performance was marginally worse than chance, $t(9) = 2.25, p = .051$, indicating that our test stimuli were biased against a positive effect of learning. The experimental group's performance was significantly different from chance, $t(9) = 7.45, p < .001$.

Participants' verbal reports in Experiments 1B and 1C were similar to those in Experiment 1A. Namely, the most common report given was that participants were basing their classification decisions on how similar or familiar the sequences were relative to the training items. None of the participants was able to verbalize any of the rules governing the sequences. Therefore, it appears that participants generally did not benefit from explicit knowledge of the sequence structure.

These results indicate that both the visual and the auditory experimental groups significantly outperformed the control groups, with participants unable to verbalize how the legal and illegal sequences differed. Hence, participants appear to have implicitly learned aspects of the statistical structure of the visual and auditory input. These initial analyses suggest commonalities among tactile, visual, and auditory statistical learning.

However, one striking difference is that the auditory test performance was substantially better than tactile or visual performance (75% vs. 62% and 63%; see Figure 2.6). Submitting these three test performances to an ANOVA reveals a main effect of modality, $F(2, 27) = 3.43, p < 0.05$, with the effect due to the auditory performance being significantly better than both touch and vision ($ps < .05$). Thus, it appears that in this task, auditory statistical learning was more proficient than both tactile and visual learning. This is in accord with

previous research emphasizing audition as being superior among the senses in regard to temporal processing tasks in general (e.g., Friedes, 1974; Handel & Buffardi, 1969; Sherrick & Cholewiak, 1986).

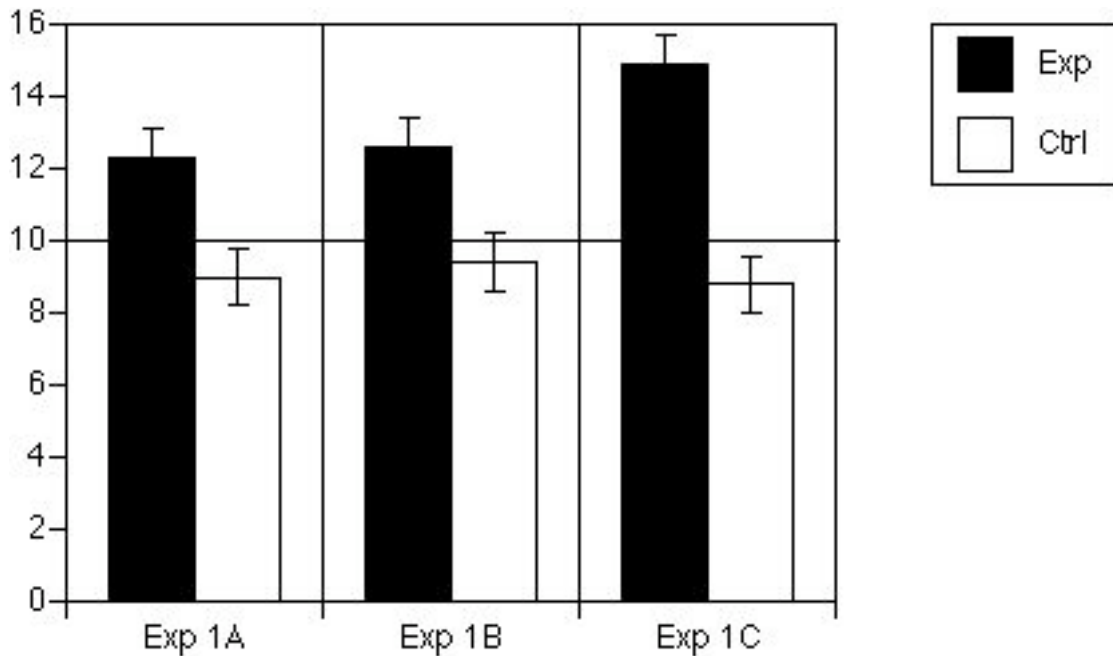


Figure 2.6. Experiment 1: Mean number of correct test responses out of 20 (plus standard error) for the experimental (indicated by solid bars) and control (indicated by open bars) groups. Ten is the level expected for chance performance.

Discussion

The previous analyses have offered a quantitative comparison among tactile, visual, and auditory learning, revealing better learning in the auditory condition. One possible objection to this conclusion is that the auditory experiment differs from the first two experiments in that pitch, instead of space, is the primary stimulus dimension being manipulated. A different possibility would have been to set up five speakers at five different spatial locations, each one producing the same pitch stimulus at different times in the

sequence, much like the visual stimuli were displayed in Experiment 1B. However, it has been proposed that for the auditory modality, pitch is, in a sense, equivalent to space (Kubovy, 1988). Shamma (2001) argued that the auditory nervous system transforms sound input, through the cochlea, into spatiotemporal response patterns, and therefore the visual and auditory systems process spatial and temporal input, respectively, in computationally similar ways. Thus, the perception of pitch and the perception of visual-spatial patterns may arise through similar computational algorithms in the two sensory modalities. For this reason, we believe that the most appropriate test for auditory statistical learning is to use stimulus elements that differ along the dimension of pitch rather than that of space. This is consistent with previous tests of auditory AGL to use stimulus elements that vary in terms of pitch or syllable rather than space. Although this research has found similar statistical learning performances in vision and audition (Fiser & Aslin, 2002; Saffran, 2002), our data suggest a quantitative advantage for auditory learning relative to tactile and visual learning.

We might also ask whether there were any *qualitative* learning differences among the three modalities. For example, were there particular test sequences within each modality that participants were better or worse at correctly endorsing? Which types of statistical information did participants within each modality rely on to perform the test task? To answer these questions, we present several additional analyses.

We first investigated whether certain sequences were easier or more difficult to classify for each modality. We conducted item analyses across the three sense modalities, entering the test performance data averaged across subjects for each sequence. This two-way ANOVA (Modality \times Sequence)

resulted in main effects of modality, $F(2, 540) = 4.73, p < .01$, and sequence, $F(19, 540) = 1.69, p < .05$, but no interaction of modality and sequence, $F(38, 540) = 1.20, p = .2$.

To get a better idea about which sources of information are most valuable for each modality, we analyzed each test sequence in terms of the information content that participants may have used to guide test performance. We used five fragment measures: associative chunk strength, novelty, novel fragment position (NFP), initial anchor strength (I-anchor), and final anchor strength (F-anchor). Associative chunk strength is calculated as the average frequency of occurrence of each test item's fragments (bigrams and trigrams), relative to the training items (Knowlton & Squire, 1994). Novelty is the number of fragments that did not appear in any training item (Redington & Chater, 1996). NFP is measured as the number of fragments that occur in novel absolute positions where they did not occur in any training item (Johnstone & Shanks, 1999). We designed the I-anchor and F-anchor measures to indicate the relative frequencies of initial and final fragments in similar positions in the training items. Previous studies used a single anchor strength measure (e.g., Knowlton & Squire, 1994) instead of calculating the initial and final measures separately, as we do here. We consider I-anchor and F-anchor separately to determine whether modality constraints lead participants to be more or less sensitive to the beginnings or endings of sequences.⁷ Finally, we used a measure of global similarity, which is the number of elements by which a test item is different from the nearest training item (Vokey & Brooks, 1992).

⁷ Meulemans and Van der Linden (2003) also used separate I-anchor and F-anchor measures.

We computed these six measures for each of the 20 test sequences, and the results are listed in Table 2.2. Inspection of this table reveals that the legal and illegal test sequences differ considerably in terms of their chunk, I-anchor, F-anchor, novel, and NFP information. It is therefore likely that one or more of these information sources guided participants in making their classification judgments at test.

To see which information sources were used for each modality, we used regression analyses. Our initial regression model contained the six sources of information listed in Table 2.2 as predictors, in addition to two other predictors: length of each sequence, as measured by the number of elements per sequence, and legality, which was simply an index of whether the sequence was legal or illegal. Because these eight predictors are highly correlated with one another, we submitted them to a principal-components analysis (PCA) to reduce the number of predictors to use in the regression analyses. The results of the PCA revealed that the eight predictors could be reduced to two components, explaining 87.7% of the variance. These two components are listed in Table 2.3.

Table 2.3

Results of Principal Component Analysis

	Component 1	Component 2
Chunk	0.950	0.179
Novel	-0.953	-0.002
NFP	-0.945	0.153
Sim	-0.696	0.536
Length	-0.154	0.949
I-Anchor	0.903	0.007
F-Anchor	0.846	0.302
Legality	0.947	0.184

Note. NFP = Novel fragment position; Sim = similarity; I-anchor = initial anchor strength; F-anchor = final anc. str.

As can be seen, the first component is roughly a measure of chunk strength, including I-anchor and F-anchor, and is also an inverse measure of novelty and NFP. This is intuitive, because a sequence with a high chunk or anchor strength contains fewer novel fragments. The second component is nearly equivalent with length. With these results in mind, we decided to use three predictors in our multiple regression model: I-anchor, F-anchor, and length. Note that in essence, what we did was separate the first component (which is roughly equivalent to chunk strength) into initial and final chunk strength predictors. We did this with the expectation that the multiple regression analysis might reveal possible modality constraints related to beginning or ending sequence biases.

The results of the regression analyses will inform us as to which of these three measures best predict whether a participant in each sensory condition will endorse a test sequence. We performed one linear regression for each modality. The results reveal that length ($p < .05$) and I-anchor ($p < .005$) were good predictors for tactile endorsements. F-anchor ($p < .005$) was a good predictor for auditory endorsements. None of the three predictors was a statistically significant predictor for visual endorsements.

In summary, the item analyses revealed no differences in terms of performance on individual sequences across the modalities. However, the multiple regression analyses revealed that there may be differences in terms of which sources of information are most important for test performance in each of the three modalities. We found that tactile learners were most sensitive to the length of the sequence and the fragment information at the beginning of a sequence, auditory learners were most sensitive to fragment information at the end of a sequence, and visual learners were biased toward neither the beginning nor the ending of the sequences. Thus, these preliminary analyses suggest that not only does auditory statistical learning of tone sequences have a quantitative advantage over tactile and visual learning, there also may be qualitative differences among the three modalities. Specifically, tactile learning appears to be sensitive to initial item chunk information, whereas auditory learning is most sensitive to final item chunk information.

Experiment 2: Tactile, Visual, and Auditory Statistical Learning

The first three experiments assessed statistical learning of tactile, visual, and auditory sequences. The results suggest the presence of modality differences affecting learning. Specifically, there was a *quantitative* learning

difference in that auditory learning was superior to the other two senses. There was also evidence for *qualitative* learning differences in that the sense modalities appeared to be differentially sensitive to the initial or final aspects of the sequences. However, one unresolved question is whether the observed learning differences are merely the result of low-level, perceptual effects of the particular stimulus elements used in the three experiments. For example, it is possible that auditory learning was more effective because the set of tones used in Experiment 1C may have been more distinctive than the set of vibration pulses or visual stimuli used in Experiments 1A and 1B. Similarly, recall that auditory training performance was significantly better than visual or tactile performances; perhaps the superior auditory test scores were due to better performance in the training phase.

To better control for perceptual and training effects, we conducted Experiment 2, which was similar to the first set of experiments except for several crucial modifications. We used a pretraining phase to assess the perceptual comparability of the stimulus elements across modalities. Also, we used a modified training task in which participants observed a sequence followed by a bigram fragment and then judged whether the bigram fragment had occurred within the sequence. We adopted this new training task to ensure similar training performance levels across the three modalities. In addition, we used a randomized design to ensure that any differences across conditions were not the result of differences in population samples. Finally, we provided a more substantive test for qualitative learning differences by incorporating test stimuli that could better assess whether participants were differentially sensitive to statistical information in the beginnings or endings of sequences. Our hypothesis, following the analyses of Experiment 1, was

that participants would be more sensitive to the initial fragments when exposed to tactile sequences, whereas they would be more sensitive to the final fragments when exposed to auditory sequences.

Method

Participants

An additional 48 undergraduates (8 in each condition) were recruited from introductory psychology classes at Cornell University.

Apparatus

The apparatus was the same as in Experiment 1.

Materials

To generate the stimuli used for Experiment 2, we created a new finite-state grammar (Figure 2.7). This grammar was created with two main constraints in mind. First, we intended it to be more complex than that used in Experiment 1. The new grammar can generate up to 75 sequences between three and seven elements in length (as opposed to 23 sequences in Experiment 1), allowing for a more difficult learning task. Second, we created the new finite-state grammar to allow us to test the hypothesis that learners are more or less sensitive to beginning or ending aspects of sequences in each sense modality. The grammar is symmetrical in terms of the number of possible bigrams and trigrams allowed in initial and final positions.⁸ Thus, it is not biased toward the beginning or ending aspects of sequences in terms of the amount of chunk information available. This allows us to have better control over what parts of the sequences may be useful for the learner.

⁸ There are 6 unique initial bigrams, 6 unique final bigrams, 13 unique initial trigrams, and 13 unique final trigrams.

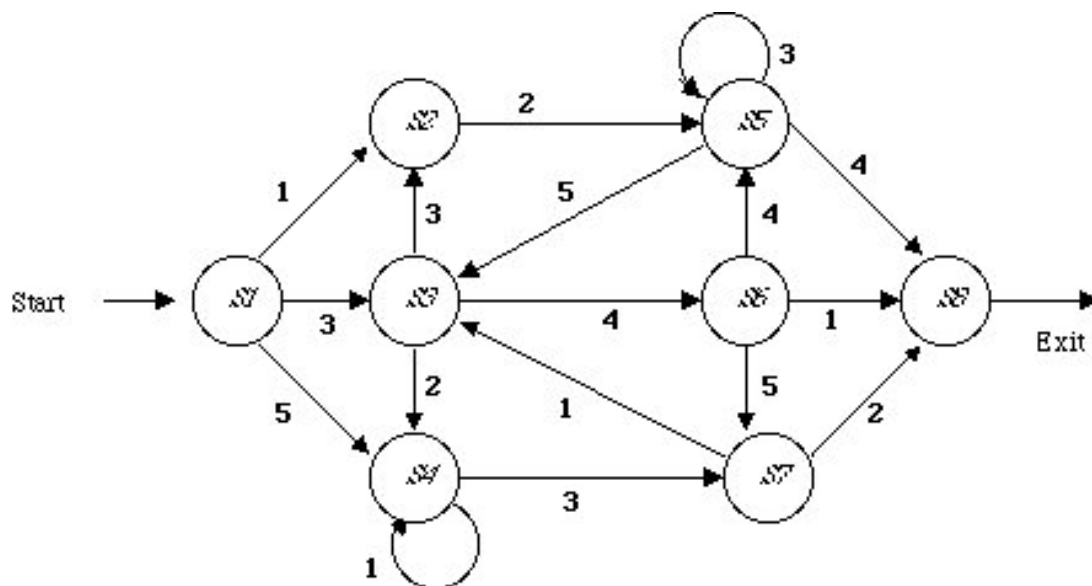


Figure 2.7. Artificial grammar used in Experiment 2.

The five stimulus elements making up the sequences were identical to those used in Experiment 1 except for the auditory tones. The tone set used for the auditory stimuli was slightly different from before, consisting of 220 Hz, 246.9 Hz, 261.6 Hz, 277.2 Hz, and 329.6 Hz (i.e., the musical notes A, B, C, C#, and E, respectively). As with the previous tone set, we used these tones because they avoid familiar melodies (Dowling, 1991). Additionally, this new tone set spans a smaller frequency range (220 Hz to 329.6 Hz, as opposed to 261.6 Hz to 493.8 Hz).

We also tested all materials for their discriminability across modalities. Ten separate participants took part in a discrimination task in which they received two stimuli (within the same modality) and judged whether they were the same or different. Participants were presented with all of the possible pairwise combinations for each modality. The data revealed that participants

were able to correctly discriminate the stimuli at near-perfect levels across all three modalities (tactile: 95%; visual: 98.3%; auditory: 98.8%), with no statistical difference in performance among modalities ($p = .87$).

Pretraining phase. For the pretraining phase, each of the five stimulus elements was paired with each other to give every possible combination ($5^2 = 25$ possible combinations). Because responses for pairs such as 3–2/2–3 and 1–4/4–1 were averaged together in the analysis (see the *Results* section), we presented the 5 pairs that contain identical elements two times instead of once (e.g., 1–1, 2–2). This gave a total of 30 stimulus pairs. Each stimulus element had a duration of 250 ms, and elements were separated by 250 ms. The pretraining materials are listed in Table 2.4.

Table 2.4

Pre-training Materials for Experiment 2

1-1 (x2)	1-2	1-3	1-4	1-5
2-1	2-2 (x2)	2-3	2-4	2-5
3-1	3-2	3-3 (x2)	3-4	3-5
4-1	4-2	4-3	4-4 (x2)	4-5
5-1	5-2	5-3	5-4	5-5 (x2)

Note. (x2) indicates that this element pair was presented twice.

Training phase. A total of 24 legal sequences were generated from the new finite-state grammar and used for the training phase. Each of these sequences was coupled with a particular bigram fragment. For half of the sequences, the bigram appeared within the sequence (e.g., 3–4–5–1–2–3–2 and 1–2). For the other half of the sequences, the bigram itself did not occur within

the sequence, but the elements composing the bigram did (e.g., 1–2–3–5–2–3–2 and 1–3). In all cases, the bigrams presented after the sequence were legal according to the finite-state grammar. Each stimulus element had a duration of 250 ms and was separated from the elements before and after by 250 ms. A 2-s pause separated the sequence from the bigram. The training materials are listed in Table 2.5.

Table 2.5

Training Materials for Experiment 2

5-3-1-3-2-3-4 / 5-3	3-4-4-5-4-4-4 / 3-5
3-2-3-1-4-5-2 / 2-3	3-2-3-1-3-2-4 / 3-4
1-2-3-5-4-4-4 / 3-5	3-4-4-5-4-5-2 / 5-3
3-4-5-1-4-5-2 / 1-4	5-1-3-1-2-3-2 / 2-1
3-4-5-1-2-3-2 / 1-2	3-3-2-5-4-5-2 / 4-3
3-3-2-5-2-3-2 / 2-3	3-2-3-1-4-4-4 / 4-1
5-1-1-1-1-3-2 / 3-2	1-2-3-5-2-3-2 / 1-3
3-4-5-1-4-4-4 / 3-4	5-3-1-4-4-3-4 / 4-5
5-3-1-4-5-2 / 3-1	3-4-5-1-4-1 / 4-3
3-4-4-5-4-1 / 4-5	5-1-3-1-4-1 / 4-3
5-3-1-4-4-4 / 4-4	3-4-4-3-3-4 / 3-2
3-2-1-1-3-2 / 3-2	5-3-1-2-3-2 / 5-1

Note. The numbers refer to a particular finger vibration, visual stimulus, or auditory tone.

Test phase. The test set consisted of 16 novel legal and 16 novel illegal sequences. Legal sequences were produced from the finite-state grammar in the normal fashion. We produced illegal sequences by changing two elements

of each legal test sequence. We created 8 of the illegal sequences, referred to as illegal-initial sequences, by modifying the second and third elements of a legal sequence (e.g., legal: 5-1-3-1-4-5-2; illegal: 5-5-2-1-4-5-2). We created the other 8 illegal sequences, referred to as illegal-final sequences, by modifying the third-to-last and second-to-last elements of a legal sequence (e.g., legal: 3-2-3-1-2-3-2; illegal: 3-2-3-1-5-2-2). Each illegal sequence was paired with the legal sequence from which it was generated, counterbalanced so that all sequences appeared both first and last, giving a total of 32 test pairs. Each stimulus element had a duration of 250 ms and was separated by 250 ms. A 2-s pause separated one sequence from the next within a pair. Table 2.6 lists the test materials.

We created the Experiment 2 test sequences so that information about legal element repetitions would not be useful. For instance, Table 2.6 reveals that out of the 32 test sequences, 18 are relevant for element repetitions, and the other 14 sequences are neutral in regard to element repetition information. If one uses the strategy of choosing the sequence within a pair containing legal element repetitions (i.e., those repetitions seen in the training sequences), this would lead to only 8 out of 18 correct endorsements. Thus, such a strategy is actually worse than random guessing, meaning that the test sequences are well controlled in terms of element repetition information.

Additionally, as we did in Experiment 1, we can analyze the test sequences in terms of chunk, novelty, and similarity information in relation to the training set. We divided the test set into four groups: legal-initial, illegal-initial, legal-final, and illegal-final. We then analyzed each group in terms of the fragment measures and made statistical comparisons among the various groups.

Table 2.6

Fragment Measures for Experiment 2 Test Sequences

Item	Chunk	Legal-Initial Sequences				
		Novel	NFP	Sim	I-Anchor	F-Anchor
5-1-3-1-4-5-2	6.27	0	0	2	2.50	5.00
1-2-3-3-5-4-1	3.27	2	8	3	2.00	2.00
3-3-2-3-3-4	5.56	1	3	2	2.00	2.00
1-2-5-3-2-3-4	4.91	3	5	3	1.00	2.00
3-4-5-1-3-2-4	6.09	0	1	2	6.00	1.00
1-2-5-4-4-3-4	4.18	1	5	3	1.00	2.00
1-2-5-4-4-4	5.56	1	5	3	1.00	5.00
1-2-5-3-2-4	3.22	3	7	4	1.00	1.00
Averages:	4.88	1.38	4.25	2.75	2.06	2.50

Table 2.6 (Continued)

Illegal-Initial Sequences						
Item	Chunk	Novel NFP		Sim	I-anchor	F-anchor
5-5-2-1-4-5-2	4.00	4	6	3	0.00	5.00
1-1-1-3-5-4-1	2.73	1	9	5	0.00	2.00
3-5-5-3-3-4	2.44	4	6	2	0.00	2.00
1-4-4-3-2-3-4	6.73	1	3	3	0.00	2.00
3-1-2-1-3-2-4	4.09	2	6	2	0.00	1.00
1-1-1-4-4-3-4	4.91	1	3	2	0.00	2.00
1-5-1-4-4-4	6.44	2	4	2	0.00	5.00
1-4-2-3-2-4	4.67	3	8	3	0.00	1.00
Averages:	4.50	2.25	5.62	2.75	0.00	2.5
Legal-Final Sequences						
1-2-3-3-3-4	4.33	2	3	3	2.00	2.00
3-2-3-1-2-3-2	7.82	0	0	2	3.50	6.00
3-3-2-3-3-3-4	4.82	2	6	3	2.00	2.00
3-4-4-3-3-3-4	4.82	1	3	3	6.00	2.00
3-2-1-3-1-4-1	5.00	1	7	4	2.50	2.50
3-4-4-3-5-4-1	4.18	1	6	3	6.00	2.00
3-2-3-1-4-1	7.00	0	0	2	3.50	2.50
1-2-3-5-3-2-4	4.27	2	3	2	2.00	1.00
Averages:	5.28	1.12	3.50	2.75	3.44	2.50

Table 2.6 (Continued)

Item	Chunk	Illegal-Final Sequences				
		Novel	NFP	Sim	I-anchor	F-anchor
1-2-3-1-1-4	5.22	2	3	3	2.00	0.00
3-2-3-1-5-2-2	4.55	5	5	2	3.50	0.00
3-3-2-3-5-4-4	5.64	0	7	4	2.00	2.50
3-4-4-3-5-5-4	3.73	4	6	3	6.00	0.00
3-2-1-3-4-3-1	4.18	4	6	4	2.50	0.00
3-4-4-3-2-1-1	5.09	1	5	4	6.00	0.00
3-2-3-5-1-1	5.11	1	4	3	3.50	0.00
1-2-3-5-4-1-4	4.00	1	2	1	2.00	0.00
Averages:	4.69	2.25	4.75	3.00	3.44	0.31

Table 2.6 shows the associative chunk strength, I-anchor, F-anchor, novelty, NFP, and similarity measures for each of these four groups. Legal-initial and illegal-initial items differed only in terms of I-anchor (2.06 vs. 0.00, $p < .05$). Likewise, legal-final and illegal-final items differed only in terms of F-anchor (2.50 vs. 0.31, $p < .05$). Legal-initial and legal-final items were statistically identical across all measures ($ps > .2$). Illegal-initial and illegal-final items differed in terms of both I-anchor (0.00 vs. 3.44, $p < .001$) and F-anchor (2.50 vs. 0.31, $p < .05$). Thus, in terms of fragment information, the only differences among the four groups of test sequences lies among the dimensions of initial and final chunk anchor strengths. This means that we can clearly examine differences in participants' sensitivities to initial and final fragment information across the three sensory modalities.

Procedure

The overall procedure was similar to that of the previous experiments but included an extra pretraining phase as well as a modified training task. Participants were randomly assigned to one of six conditions: tactile, visual, auditory, tactile control, visual control, or auditory control. The three control conditions were identical to their respective experimental conditions except that the controls participated in the pretraining and test phases only.

All participants in the tactile conditions were assessed by the Edinburgh Handedness Inventory (Oldfield, 1971) to determine their preferred hand.

Pretraining phase. As already described, a separate group of participants had participated in a simple discrimination task, which revealed that the stimuli are easily discriminable across the modalities. To provide an additional test of perceptual comparability, we incorporated the pretraining phase into the current experiment. As an additional benefit, this procedure also served to familiarize participants with the actual stimulus elements before they were exposed to the training sequences.

Participants were informed that they would observe two stimuli, one following the other. The stimuli consisted of vibration pulses, visual stimuli, or tones, depending on the experimental condition. Participants were required to judge how similar the two stimuli were to each other and give a rating between 1 and 7, where 1 corresponded to most dissimilar and 7 to most similar. Participants in the tactile conditions were told to base their ratings on the vibration pulses' proximity to each other, as all vibration pulses were identical except for which fingers were stimulated. Similarly, participants in

the visual conditions also were told to base their ratings on the stimuli's proximity, as the stimuli themselves were identical and differed only in terms of where they were located. Participants in the auditory conditions were told to base their ratings on the pitches of the tones.

Before the rating task began, participants were exposed to each of the five possible stimuli, one at a time, so that they knew what the possible stimuli were. Then they were presented with each of the 30 possible pairs listed in Table 2.4, in random order for each participant. All stimuli were delivered for a duration of 250 ms with a 250 ms pause occurring between the stimuli within a pair. A prompt containing a reminder of the rating scheme appeared on the screen, and the participant used the keyboard to give a numerical response between 1 and 7. Following a 2-s pause after the rating was given, the next stimulus pair was delivered.

Training phase. As in Experiment 1, the purpose of the training phase was for the participants to attend to the legal training sequences without explicit instruction that the sequences contained statistical regularities. On the basis of pilot studies, we modified the training procedure slightly from Experiment 1 in an attempt to equate training performance across the three modalities.

At the beginning of the training phase, participants were instructed that they would observe a particular sequence of stimuli and then, after a slight pause, would observe two additional elements. The task was to decide whether the pair of elements had occurred within the sequence in the same order and then to press the appropriate key, *Y* for yes, *N* for no. The training sequence–pair combinations from Table 2.5 were presented in random order for three blocks, for a total of 72 training trials. Stimulus elements had a

duration of 250 ms and were separated by 250-ms pauses. A 2-s pause occurred between each sequence and each pair of elements. One second after the last element of the stimulus pair occurred, a prompt was displayed on the screen asking for the participant's response. The next sequence-pair combination began after a 2-s pause.

Test phase. The purpose of the test phase was to assess how well participants learned the statistical regularities of the training set and could generalize such knowledge to novel stimuli in a classification task. At the beginning of the test phase, participants were instructed that all of the sequences they had been exposed to in the previous phase of the experiment were generated by a complex set of rules. They now would be exposed to new sequences, presented in groups of two. One of the sequences in each pair was generated by the same rules as before, whereas the other was not. The participants' task was to choose which sequence was generated from the same rules by pressing a key marked 1 or 2, signifying the first or second sequence, respectively. The test sequence pairs from Table 2.6 were presented in random order for each participant. Stimulus elements had a duration of 250 ms and were separated by 250-ms pauses. A 2-s pause occurred between the two sequences of a pair. One second after the second sequence occurred, a prompt was displayed on the screen asking for the participant's response. The next pair of sequences began following a 2-s pause.

Results

Pretraining Phase

We collected similarity ratings from all participants and averaged them for each element pair combination to form three similarity matrices, one for

each modality. Within each modality, the experimental and control ratings were combined. We submitted each similarity matrix to a multiple dimensional scaling (MDS) procedure (euclidian model) using SPSS 10.0 for Windows. A one-dimensional solution provided a good fit for each of the three modalities, with stress values less than 0.0500 in all cases (touch = 0.0056; vision = 0.0086; audition = 0.0470). As can be seen from Figure 2.8, the tactile and visual solutions contain clearly separated stimuli, in accord with the linear relationship of the actual stimuli (e.g., the tactile vibration elements are in the expected order, beginning with the thumb pulse, then the second finger, then third finger). In slight contrast, the auditory solution contains two tones, the third (261.6 Hz) and fourth (277.2 Hz), that are clustered together in state space.

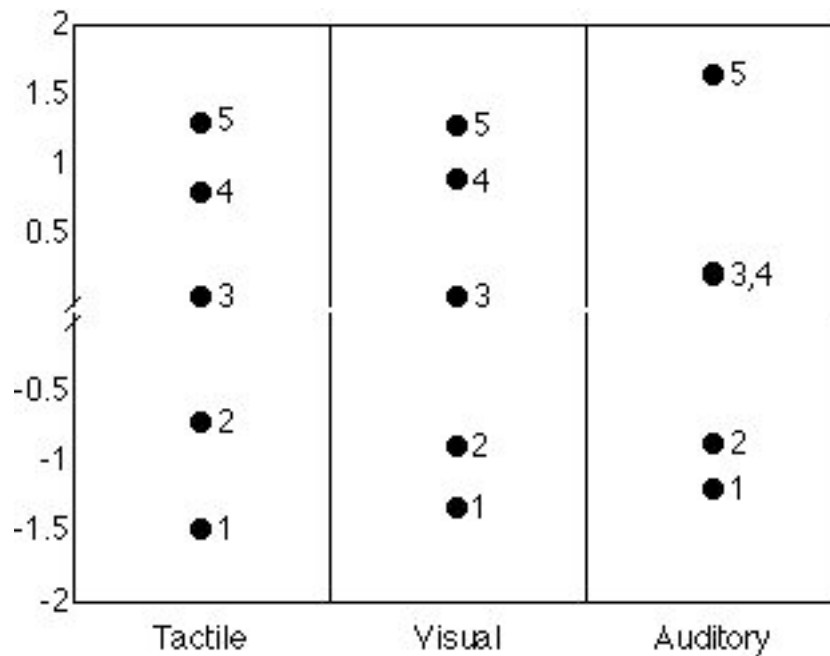


Figure 2.8. MDS solutions using the pretraining data for each of the three sensory modalities. The numbers correspond to each of the five possible stimulus elements for each modality.

We interpret these MDS solutions as depicting that, overall, the stimuli in the three modalities are perceived in psychologically similar ways. The only noticeable difference is that two of the tones may have similar perceptual representations.

Training Phase

The mean training performance out of 72 for each modality was 43.38 (60.3%) for tactile, 50.13 (69.6%) for visual, and 48.25 (67.0%) for auditory. The data were submitted to a one-way ANOVA with the factor of modality. There was a marginally significant main effect, $F(2, 21) = 2.81, p = .083$. Post hoc tests revealed a significant difference between the tactile and visual training means ($p < .05$). These results indicate that although the tactile training performance was somewhat lower than the auditory and visual performances, in general, scores across the three modalities were roughly equivalent.

Test Phase

The mean test scores out of 32 for each group were 15.75 for tactile–control (49.2%), 16.00 for tactile–experimental (50.0%), 14.38 for visual–control (44.9%), 15.88 for visual–experimental (49.6%), 15.25 for auditory–control (47.6%), and 19.25 for auditory–experimental (60.2%). Figure 2.9 displays Experiment 2 test performance.

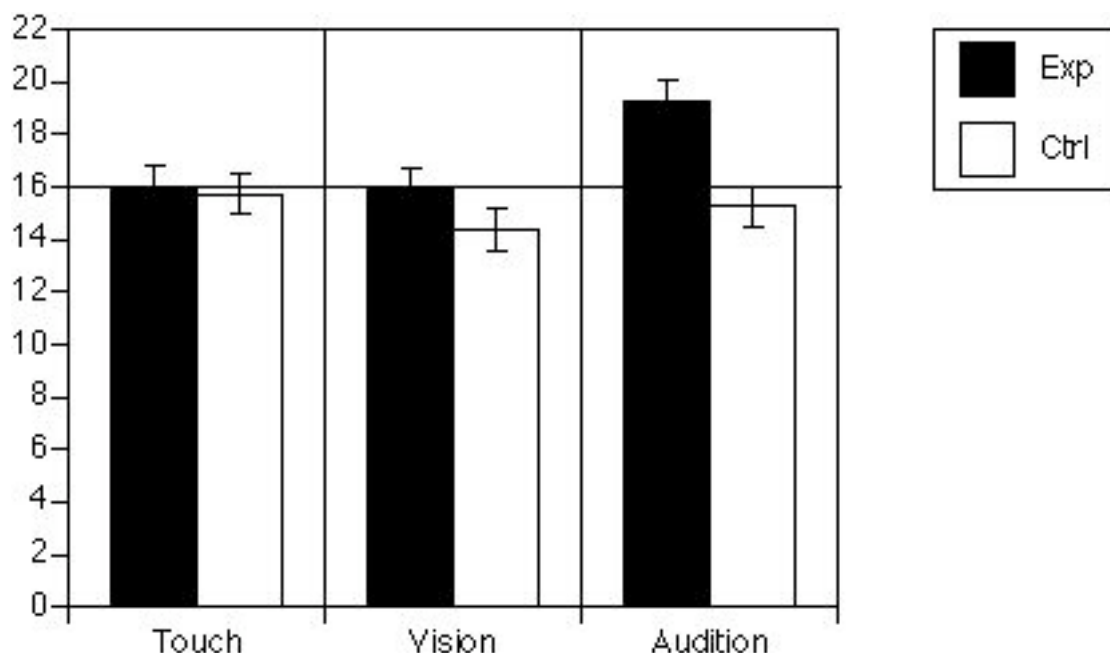


Figure 2.9. Experiment 2: Mean number of correct test responses out of 32 (plus standard error) for the experimental (solid bars) and control (open bars) groups. Sixteen is the level expected for chance performance.

Recall that the control groups participated in the pretraining and test phases only. Therefore, the best way to assess learning within each sensory modality is to compare experimental group performance with both hypothetical chance levels (i.e., 50%) and control group performance (see Redington & Chater, 1996). Individual single group t tests comparing experimental group performance with theoretical chance reveal that only the auditory group performed significantly better than chance, $t(7) = 3.61, p < .01$. Likewise, paired t tests reveal that only the auditory experimental group performed better than its control group, $t(7) = 3.86, p < .01$. These results reveal that only the auditory experimental group learned the statistical regularities of the training corpus.

We next submitted the experimental group data to a repeated-measures ANOVA with the between-subjects factor of modality and the within-subject

factor of initial versus final group sequences. We found a main effect of modality, $F(2, 21) = 4.95, p < .05$, a main effect of initial–final group, $F(2, 21) = 8.61, p < .01$, and no interaction ($p = .25$). The effect of initial–final group arises because test performance across modalities was generally better at initial group sequences. The main effect of modality arises because the auditory group was significantly better than both the tactile and the visual groups ($ps < 0.05$). Thus, these analyses confirm the presence of modality differences in learners' test performances.

The ANOVA revealed no interaction between modality and initial versus final group sequences. However, because Experiment 1 suggested the presence of such modality differences, we continued to explore possible differences with planned comparisons. As described earlier, the test sequences were created such that test pairs differed only in terms of I-anchor and F-anchor measures. Thus, we can easily determine whether each sensory modality was better at discriminating sequences on the basis of initial or final fragment information. For the experimental group test performance in each modality, we considered initial and final test pairs separately. Performance on the 16 initial test pairs (i.e., test pairs consisting of one legal–initial and one illegal–initial sequence) was 8.50 (53.1%) for touch, 9.50 (59.4%) for vision, and 10.13 (63.3%) for audition. We conducted a one-way ANOVA that revealed no statistical differences among the three modalities, $F(2, 21) = 1.69, p = .21$.

Next, we consider performance on the 16 final test pairs for the experimental conditions (i.e., test pairs consisting of one legal–final and one illegal–final sequence). Performance was 7.50 (46.9%) for tactile, 6.38 (39.8%) for visual, and 9.13 (57.0%) for auditory. An ANOVA revealed a main effect of

modality, $F(2, 21) = 3.98, p < .05$. Specifically, auditory performance was significantly better than visual performance ($p < .05$).

To summarize, the auditory group alone showed learning. Furthermore, the auditory superiority appears to be largely due to better performance, relative to vision and touch, on fragment–final test pairs. Thus, consistent with Experiment 1, Experiment 2 provides evidence for both quantitative and qualitative learning differences across the senses.

Discussion

Experiment 2 more closely examined modality differences in statistical learning by attempting to control for two variables that could have influenced performance in Experiment 1: low-level perceptual factors and training performance effects. We controlled the first variable by introducing a pretraining phase, in which participants observed all combinations of element pairs and gave similarity ratings for each. These ratings were then submitted to an MDS analysis, and the results provide good indication that, across modalities, the stimuli were represented similarly. Furthermore, an additional control task revealed that the stimuli within each modality were easily discriminated.

Even if the stimuli across all three conditions were comparable in terms of their perceptibility, it is still possible that uneven training performances could lead to differences in test performance. For example, in Experiment 1, the auditory group's training performance was significantly better than both visual and tactile training performance. The improved training performance may have led to better encoding of the relevant fragment information, resulting in a better ability to classify novel sequences at test. For Experiment

2, we controlled training performance effects by using a training task that resulted in relatively comparable training scores. This task proved more difficult than that used in the first set of experiments, with scores ranging between 60% and 69%. Auditory training performance was equivalent to visual performance and only slightly better than tactile performance.

Even after we controlled for the perceptual and training effects, auditory test performance still was significantly better than tactile and visual performance. In fact, only the auditory group showed a main effect of learning. Experiment 2 also shows that the auditory modality's better performance was due to a heightened sensitivity to sequence-final fragment information.

Comparing the results of Experiment 2 with those of Experiment 1 reveals strong similarities. Both experiments showed an auditory advantage for classifying novel sequences in regard to their legality. Both experiments also revealed differences in terms of whether initial or final fragment information was more important for each modality. The regression analyses in the first set of experiments suggested a tactile-initial and an auditory-final effect. The auditory-final effect was confirmed in Experiment 2.

One potentially troubling aspect of Experiment 2 is that there appeared to be no tactile learning. This may be a result of the more complex grammar and more difficult test that we used in Experiment 2. Tactile memory for serial lists may be weaker compared with vision and audition (Mahrer & Miles, 1999) and, if so, may hinder learning during the training phase. The lack of learning leaves open the possibility that the tactile learning we found in

Experiment 1A was spurious. To replicate the tactile learning from Experiment 1A, we conducted an additional tactile learning experiment.⁹

We hypothesized that using a slightly modified training scheme, one that is less computationally demanding, might allow participants to better encode the training regularities and lead to successful learning. Instead of a training sequence being followed by a pair of elements (i.e., a bigram), each training sequence was followed by a single element. The participants' task was to judge whether the single element had been a part of the preceding sequence. In all other respects this additional experiment was identical to the tactile condition in Experiment 2. On the training task, participants scored 62.2 out of 72 (86.4%), and on the test task they scored 18.63 out of 32 (58.2%). An ANOVA (experimental vs. control) comparing the test performance with the control performance revealed a significant effect, $F(1, 14) = 5.83, p < .05$. Thus, under a slightly less complex training condition that allowed better training performance, tactile participants showed a main effect of learning in the test phase. This confirms the tactile learning we found in Experiment 1.

Finally, it could be argued that the superior performance in the auditory condition was due to the presence of relative pitch information present in the tone sequences (e.g., see Saffran & Griepentrog, 2001). Because the auditory sequences can be construed as melodies, they contain information about not just absolute pitch but also relative pitch in the form of familiar musical intervals (e.g., perfect fourths and major sevenths). To eliminate this information and provide an auditory task more comparable to the other experiments, we conducted a new auditory experiment identical to

⁹ Although there also was no effect of learning in Experiment 2's visual condition, we did not feel it was necessary to conduct an additional visual learning experiment because visual statistical learning has been demonstrated in previous research (e.g., Fiser & Aslin, 2002).

Experiment 2 but with a different tone set. The new tones were created so that they would neither conform to standard musical notes nor contain standard musical intervals between them. The tones, having frequencies of 210 Hz, 245 Hz, 286 Hz, 333 Hz, and 389 Hz, were equally spaced in log frequency, on the basis of a ratio of $7/6$ (1.166) and its powers 2, 3, and 4 (1.36, 1.59, 1.85, respectively). Of these ratios, only one comes close to a musical interval, minor 6 (ratio 1.6), which is not a very prominent interval. With this new tone set, participants scored 41.8 out of 72 (58.1%) on the training task and 18.7 out of 32 (58.4%) at test. The test scores were significantly greater than chance, $t(9) = 3.69$, $p < .01$, revealing an effect of learning. We also compared the test scores with the original Experiment 2 auditory scores and found no difference, $t(7) = 0.35$, $p = .74$. Thus, this additional experiment indicates that the presence or absence of musical interval information does not appear to affect learning and therefore is not the underlying cause for the superior auditory performance.

General Discussion

Our experiments provided a controlled investigation into the nature of statistical learning across the three primary sensory modalities. Unlike in previous statistical learning forays, we used comparable materials and identical procedures across the three senses. Additionally, in Experiment 2, we controlled for low-level perceptual as well as training performance effects, which allowed us to make direct comparisons among the modalities. In this section we discuss the main findings of these experiments and conclude by considering the underlying neural and cognitive mechanisms.

The first important finding is that touch can mediate statistical learning of sequential input. Experiment 1A revealed that after very brief (10 min)

exposure to training sequences produced from an artificial grammar, participants were able to classify novel sequences as being either generated or not generated from the same grammar. To our knowledge, no other studies have demonstrated such a tactile learning ability, perhaps because of the belief that deep structure learning is beyond the tactile sense's capacity. Participants did not perform above chance or control level performance in Experiment 2, likely because the grammar was too complex and the distinction between legal and illegal sequences more subtle. However, when we used a slightly less demanding training scheme, participants displayed learning, which confirms that touch can mediate the learning of an artificial grammar.

The tactile learning in both experiments occurred in the absence of feedback and apparently without participants' awareness. Additionally, participants learned more than simply element frequencies, sequence lengths, or beginning and ending element legality, as these factors were controlled; rather, the legal and illegal test sequences differed in terms of adjacent element statistics. The learning process observed here in the tactile domain is conceptually similar to statistical learning found in the visual and auditory modalities (e.g., Fiser & Aslin, 2002; Saffran et al., 1999). Besides having theoretical importance, these data may have practical implications in the realm of communication devices for the visually or auditorily impaired, by capitalizing on people's ability to encode and represent tactile temporal order statistics.

The second important finding is the presence of modality constraints affecting statistical learning. These modality constraints took two main forms: a quantitative effect, and a qualitative effect. The quantitative effect was evidenced by the fact that auditory statistical learning was better than both

tactile and visual learning. The auditory advantage occurred even after we controlled for training performance and stimulus element perceptibility across the modalities and eliminated standard musical intervals from the tone sequences. Our results thus suggest that the modality constraints affected learning itself rather than a lower level perceptual process that might have in turn affected learning performance. Although previous research has pointed to an auditory advantage for low-level temporal processing tasks (Mahar et al., 1994; Sherrick & Cholewiak, 1986), our results appear to be some of the first evidence that such an advantage extends to more complex processes, namely statistical learning of sequential input.

In addition to the quantitative effect was a qualitative learning effect that took the form of biases regarding which aspects of a sequence each sensory system is more or less attuned toward. We found evidence that learners of tactile sequences may have been more sensitive to fragment–initial information, whereas learners of auditory material were more sensitive to fragment–final information. These biases suggest that each sensory system may apply slightly different computational strategies when processing sequential input. The auditory–final bias is interesting because it mirrors the modality effect in serial recall, in which a more pronounced recency effect (i.e., greater memory for items at the end of a list) is obtained with spoken as compared with written lists (e.g., Crowder, 1986). This may indicate that similar constraints affect both explicit encoding of serial material and implicit learning of statistical structure. In both cases, learners appear to be more sensitive to material at the end of sequences or lists for auditory input. It may prove fruitful to further explore this hypothetical connection between serial

list learning and implicit sequential learning; the results may inform research in both domains, which have traditionally remained separate endeavors.

In contrast to the quantitative and qualitative effects we found, two previous studies comparing auditory and visual statistical learning did not report any modality differences. Fiser and Aslin (2002) found human visual statistical learning of spatiotemporal sequences to be very similar to learning in the auditory domain with temporal sequences (e.g., Aslin, Saffran, & Newport, 1998) and in the visual domain with spatial structures (e.g., Fiser & Aslin, 2001). Similarly, Kirkham et al. (2002) concluded that infant visual statistical learning was similar to auditory learning (e.g., Saffran et al., 1996). However, Fiser and Aslin's (2002) and Kirkham et al.'s (2002) conclusions were based on studies that did not use comparable stimuli or procedures across the modalities. Thus, neither of these two studies are adequate for making fine-tuned cross-modal comparisons. We agree that there are similarities in how infants and adults learn sequential patterns across vision and audition, but our data reveal important differences, as well. Saffran (2002) used more comparable procedures and stimuli in her visual and auditory AGL experiments and found comparable overall learning performances across modalities. However, a subtle modality difference was revealed, showing that auditory learners were more sensitive to the presence or absence of predictive dependencies in the grammar, perhaps because they had more experience in the auditory domain for tracking such sequential structure.

Similarly, it could be argued that the auditory learning advantage we observed was merely due to participants having more prior experience listening to tone sequences compared with feeling vibration pulses or tracking visual spatiotemporal patterns. However, if this was the case, one would

expect the auditory training performance levels to be substantially better than visual or tactile scores. In fact, auditory training performance was lower than visual performance in Experiment 2. Additionally, it is difficult to see how previous training exposure would lead to the qualitative differences we observed in terms of each sensory modality being differentially biased toward the beginning or ending of sequences.

If, as we suggest, modality constraints exist, what might be their nature? The answer clearly depends on how one views statistical or implicit learning itself. Because most researchers have tended to emphasize the similarities existing between statistical learning in various domains, this has led to implicit assumptions—or, in some cases, explicit statements (Kirkham et al., 2002)—depicting statistical learning as a single, domain-general mechanism. If this view is adopted, modality constraints are seen as influencing the processing of input sequences before the information is funneled into the presumed amodal statistical learning mechanism, allowing some types of input to be processed more or less efficiently.

Another view of statistical learning is that it is made up of various subsystems, each operating over different types of input and subserved by different brain areas. This view is supported by increasing evidence that unimodal brain areas contribute to the learning of statistical patterns. For example, in a functional magnetic resonance imaging study (Reber, Stark, & Squire, 1998), subjects learned to categorize visual patterns in terms of whether the patterns were similar to a previously seen corpus. It was found that occipital cortex (V1, V2) was instrumental for learning the structural regularities of the patterns, apart from being involved in visual perception itself. Similarly, Molchan, Sunderland, McIntosh, Herscovitch, and Schreuers

(1994) found the primary auditory cortex to be involved in auditory associative learning. These studies indicate that relatively low-level unimodal sensory areas play an important role in learning environmental statistical regularities. It is our view that implicit statistical learning may be akin to perceptual priming, where modality-specific brain areas mediate learning by becoming more fluent at processing previously observed stimuli and/or stimuli that contain similar statistical properties as those viewed earlier (also see Chang & Knowlton, 2004). Within this framework, modality constraints may reflect general processing differences that exist among the various statistical learning subsystems, with the auditory system excelling at encoding statistical relations among temporal elements and the visual system specializing primarily in computing spatial relationships.

In conclusion, we have presented new evidence revealing both similarities and differences in statistical learning across the senses. An important target for future research is to uncover to what extent these modality constraints are related to previous reports of modality differences in perception and cognition. We anticipate that future studies, involving a combination of cognitive and neurophysiological methods, will further illuminate the nature of modality-constrained statistical learning across the senses.

References

- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9, 321–324.
- Beaman, C. P. (2002). Inverting the modality effect in serial recall. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 55A, 371–389.
- Chang, G. Y., & Knowlton, B. J. (2004). Visual feature learning in artificial grammar classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 714–722.
- Cleeremans, A. (1993). *Mechanisms of implicit learning: A connectionist model of sequence processing*. Cambridge, MA: MIT Press.
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25, 257–271.
- Collier, G. L., & Logan, G. (2000). Modality differences in short-term memory for rhythms. *Memory and Cognition*, 28, 529–538.
- Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, 5, 539–546.
- Craig, J. C., & Rollman, G. B. (1999). Somesthesia. *Annual Review of Psychology*, 50, 305–331.
- Crowder, R. G. (1986). Auditory and temporal factors in the modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2, 268–278.
- Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception and Psychophysics*, 5, 365–373.

- Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics*, 50, 305–313.
- Engle, R. W., & Mobley, L. A. (1976). The modality effect: What happens in long-term memory? *Journal of Verbal Learning and Verbal Behavior*, 15, 519–527.
- Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12, 499–504.
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 458–467.
- Freides, D. (1974). Human information processing and sensory modality: Cross-modal functions, information complexity, memory, and deficit. *Psychological Bulletin*, 81, 284–310.
- Garner, W. R., & Gottwald, R. L. (1968). The perception and learning of temporal patterns. *Quarterly Journal of Experimental Psychology*, 20, 97–109.
- Gescheider, G. A. (1966). Resolving of successive clicks by the ears and skin. *Journal of Experimental Psychology*, 71, 378–381.
- Gescheider, G. A. (1967). Auditory and cutaneous perceived successiveness. *Journal of Experimental Psychology*, 73, 179–186.
- Glenberg, A. M., & Fernandez, A. (1988). Evidence for auditory temporal distinctiveness: Modality effects in order and frequency judgments. *Journal of Experimental Psychology*, 14, 728–739.
- Glenberg, A. M., & Jona, M. (1991). Temporal coding in rhythm tasks revealed by modality effects. *Memory & Cognition*, 19, 514–522.

- Glenberg, A. M., & Swanson, N. C. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 3–15.
- Gomez, R. L. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology*, 33, 154–207.
- Gomez, R. L., & Gerken, L. A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70, 109–135.
- Handel, S., & Buffardi, L. (1969). Using several modalities to perceive one temporal pattern. *Quarterly Journal of Experimental Psychology*, 21, 256–266.
- Hauser, M. D., Newport, E., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, 78, B53–B64.
- Johnstone, T., & Shanks, D. R. (1999). Two mechanisms in implicit artificial grammar learning? Comment on Meulemans and Van der Linden (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 524–531.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain-general learning mechanism. *Cognition*, 83, B35–B42.
- Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 79–91.
- Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 169–181.

- Kubovy, M. (1988). Should we resist the seductiveness of the space:time::vision:audition analogy? *Journal of Experimental Psychology: Human Perception and Performance*, 14, 318–320.
- Lechelt, E. C. (1975). Temporal numerosity discrimination: Intermodal comparisons revisited. *British Journal of Psychology*, 66, 101–108.
- Mahar, D., Mackenzie, B., & McNicol, D. (1994). Modality-specific differences in the processing of spatially, temporally, and spatiotemporally distributed information. *Perception*, 23, 1369–1386.
- Mahrer, P., & Miles, C. (1999). Memorial and strategic determinants of tactile recency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 630–643.
- Manning, S. K., Pasquali, P. E., & Smith, C. A. (1975). Effects of visual and tactual stimulus presentation on learning two-choice patterned and semi-random sequences. *Journal of Experimental Psychology: Human Learning and Memory*, 1, 736–744.
- Meulemans, T., & Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 1007–1028.
- Meulemans, T., & Van der Linden, M. (2003). Implicit learning of complex information in amnesia. *Brain and Cognition*, 52, 250–257.
- Molchan, S. E., Sunderland, T., McIntosh, A. R., Herscovitch, P., & Schreuers, B. G. (1994). A functional anatomical study of associative learning in humans. *Proceedings of the National Academy of Sciences, USA*, 91, 8122–8126.
- Oatley, K., Robertson, A., & Scanlan, P. M. (1969). Judging the order of visual stimuli. *Quarterly Journal of Experimental Psychology*, 21, 172–179.

- Oldfield, R. L. (1971). The assessment of handedness: The Edinburgh Inventory. *Neuropsychologia*, 9, 97–113.
- Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General*, 130, 401–426.
- Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory and Cognition*, 17, 398–422.
- Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, 119, 264–275.
- Pothos, E. M., & Bailey, T. M. (2000). The role of similarity in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 847–862.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855–863.
- Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford, England: 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 Science*, 95, 747–750.
- Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General*, 125, 123–138.
- Rubinstein, L., & Gruenberg, E. M. (1971). Intramodal and crossmodal sensory transfer of visual and auditory temporal patterns. *Perception and Psychophysics*, 9, 385–390.

- Saffran, J. R. (2000). Non-linguistic constraints on the acquisition of phrase structure. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the twenty-second annual conference of the Cognitive Science Society* (pp. 417–421). Hillsdale, NJ: Erlbaum.
- Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, 47, 172–196.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology*, 37, 74–85.
- 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.
- Segar, C. A. (1994). Implicit learning. *Psychological Bulletin*, 115, 163–196.
- Shamma, S. (2001). On the role of space and time in auditory processing. *Trends in Cognitive Sciences*, 5, 340–348.
- Sherrick, C. E., & Cholewiak, R. W. (1986). Cutaneous sensitivity. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance*, Vol. I: *Sensory processes and perception* (pp. 12/1–12/58). New York: Wiley.
- Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 328–344.

Wright, A. A. (2002). Monkey visual and auditory memory. In S. B Fountain, M. D. Bunsey, J. H. Danks, & M. K. McBeath (Eds.), *Animal cognition and sequential behavior: Behavioral, biological, and computational perspectives* (pp. 69–90). Boston: Kluwer Academic.

CHAPTER THREE

Seeing and Hearing in Space and Time: Effects of Modality and Presentation Rate on Implicit Statistical Learning

Intuitively, each sensory modality seems biased to handle particular aspects of environmental input. For instance, vision and audition appear to be most adept at processing spatial and temporal input, respectively (Kubovy, 1988). Consider that it takes very little time – often a single glance – in order to recognize a relatively complex visual scene (Potter, 1976) but it generally takes several seconds or longer to recognize a melody or voice. That is, a brief “snapshot” of sound is fairly incomprehensible whereas a visual snapshot may be quite coherent (Seifritz, Espoto, Neuhoff, & Di Salle, 2003). Also consider that whereas the auditory system must compute the location of sounds through differences in intensity and time of arrival at each ear, the location of visual stimuli is directly mapped onto the retina and then projected topographically into cortical areas (Bushara et al., 1999). This somewhat anecdotal evidence, bolstered by studies of perception and memory, suggests that in visual cognition, the dimensions of space weigh most heavily, whereas for audition, the temporal dimension is most prominent (Friedes, 1974; Geldard, 1970; Kubovy, 1988; O'Connor & Hermelin, 1978; Penney, 1989).

These modality constraints have been proposed to affect the manner in which stimuli are perceived (Mahar, Mackenzie, & McNicol, 1994; Repp & Penel, 2002), learned (Conway & Christiansen, 2005; Handel & Buffardi, 1969; Saffran, 2002), and maintained in working memory (Collier & Logan, 2000; Glenberg & Swanson, 1986; Penney, 1989). Even so, researchers do not always

give modality effects due attention. In fact, it has been suggested that cognitive psychologists ought not to consider the senses as separate entities at all (e.g., Marks, 1978; Stoffregen & Bardy, 2001). Or at the very least, modality effects ought to be de-emphasized, with an emphasis instead on the importance of amodal information in the environment that can be detected by any of the sense modalities (e.g., Gibson, 1966). Unfortunately, these views tend to obfuscate the possible presence of modality differences that may affect processing, which can hinder the development of a complete theory of human cognition.

It is currently unknown to what extent spatial and temporal constraints affect statistical learning, the capability for extracting environmental regularities, largely through automatic learning mechanisms operating outside of awareness. Statistical learning is a fundamental ability believed to underlie important aspects of language and cognition (Altmann, 2002; Conway & Christiansen, 2001), including speech segmentation (Saffran, Newport, & Aslin 1996) visual scene computations (Fiser & Aslin, 2001), and tactile pattern processing (Conway & Christiansen, 2005). Very few studies of statistical learning have directly compared learning across sensory domains; fewer still have considered possible spatial and temporal constraints on learning. One exception is Gomez (1997), who suggested that visual learning of artificial grammars proceeds better when the stimulus elements are presented simultaneously – that is, spatially arrayed – rather than sequentially, presumably because a simultaneous format permits better chunking of the stimulus elements. Likewise, Saffran (2002) found that participants learned predictive relationships best with an auditory or visual-simultaneous presentation, but did poorly in a visual-sequential condition.

Finally, more recently Conway and Christiansen (2005) examined statistical learning of tactile, visual, and auditory input sequences and found that auditory learning of such material exceeded both visual and tactile learning.

In this report, we extend these studies to directly test the hypothesis that temporal and spatial constraints affect visual and auditory learning of statistical patterns. Using the artificial grammar learning (AGL; Reber, 1967) paradigm, we expose participants to statistically-governed visual or auditory input sequences generated from an artificial grammar and then test learners on their ability to generalize their knowledge to novel patterns. Our primary manipulation is the manner in which the visual input is distributed: spatially, spatiotemporally, or temporally. In line with a modality-constrained view, we predict that learning will be greatest for the visual-spatial input, comparable to an auditory (temporal) condition, and poorest for the visual-temporal material. An additional aim of this study is to explore the effect that rate of presentation has on learning, an issue that has not yet been explored in AGL or traditional statistical learning tasks. Faster presentation rates may magnify the effect of modality constraints by adversely affecting learning in the non-preferred mode of processing (see Collier & Logan, 2002). Consequently, we predict that at the fast rate, learning in the visual-temporal condition will show the largest decrement in performance.

Experiment 3: Effects of Modality and Presentation Rate on Learning

In Experiment 3, we manipulated presentation format and presentation rate to uncover additional constraints on visual and auditory implicit statistical learning.

Method

Subjects

One hundred forty-four subjects (twelve in each condition) with normal hearing and normal or corrected-to-normal vision were recruited from psychology classes at Cornell University, earning course credit for their participation. Data from two subjects were discarded, one due to the participant being color blind and the other due to experimenter error.

Materials

The materials are similar to those used by Conway and Christiansen (2005). Figure 3.1 shows the artificial grammar used to generate the input sequences, taken from Gomez and Gerken (1999). The grammar consists of five unique elements, each represented by a numeral (1-5). In turn, each numeral is mapped onto elements from one of four visual or auditory vocabularies, corresponding to four input conditions (auditory, visual-temporal, visual-spatial, and visual-spatiotemporal).

As in Conway and Christiansen (2005), for the “training” phase we used 12 legal sequences that were generated from the grammar. Each legal sequence was used twice to create a set of 12 training pairs (see Table 3.1). Six of the pairs consisted of the same training sequence presented twice (matched pairs), whereas the other 6 pairs consisted of 2 different sequences (mismatched pairs). These matched and mismatched training pairs were used in conjunction with a same–different judgment task.

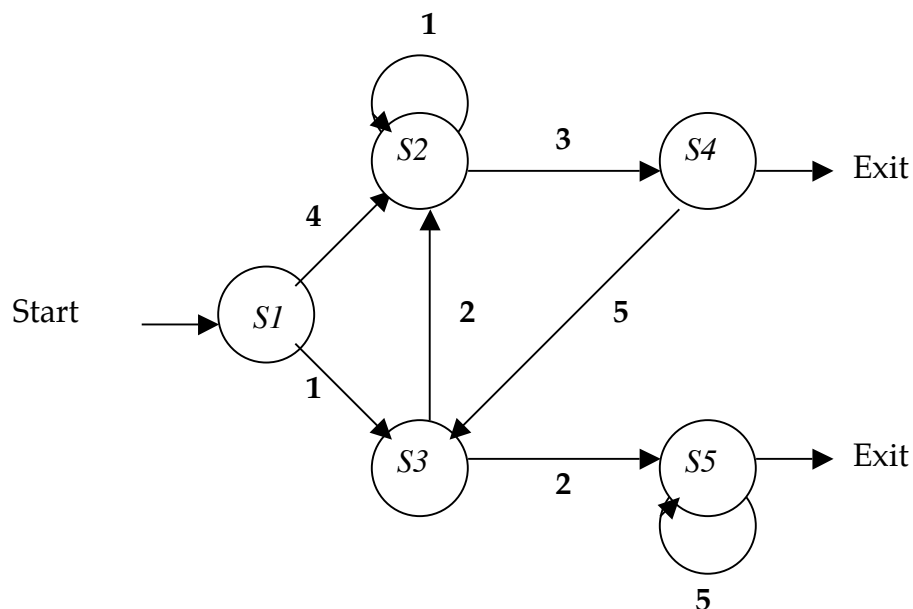


Figure 3.1. Artificial grammar used in Experiment 3, taken from Gomez and Gerken (1999). Each numeral was mapped onto a particular auditory or visual stimulus element, depending on the condition.

Table 3.1

Training Pairs

Matched Pairs	Mismatched Pairs
1-2-1-1-1-3 / 1-2-1-1-1-3	1-2-3-5-2-5 / 1-2-3-5-2-3
4-1-1-3-5-2 / 4-1-1-3-5-2	1-2-3-5-2-3 / 1-2-3-5-2-5
4-1-3-5-2 / 4-1-3-5-2	4-3-5-2-3 / 4-3-5-2-5
1-2-5-5-5 / 1-2-5-5-5	4-3-5-2-5 / 4-3-5-2-3
4-1-3 / 4-1-3	1-2-5-5 / 1-2-1-3
1-2-3 / 1-2-3	1-2-1-3 / 1-2-5-5

Note. The numbers refer to a particular visual or auditory stimulus (see text).

The test set consisted of 10 novel legal and 10 illegal sequences. Legal sequences were produced from the finite-state grammar in the normal fashion. The illegal sequences each begin with a legal element (i.e., 1 or 4), followed by one or more illegal transitions and ending with a legal element (i.e., 2, 3, or 5). For example, the illegal sequence 1-4-5-1-3-3 begins and ends with legal elements (1 and 3, respectively) but contains several illegal interior transitions not allowed by the grammar (1-4, 4-5, 5-1, and 3-3). The legal and illegal sequences can be described as differing from one another in terms of co-occurrence statistics of adjacent elements. That is, a statistical learning mechanism could discern which novel test sequences are illegal by noting the presence of pairwise element combinations that did not occur in the training set. All test sequences are listed in Table 3.2.

As mentioned earlier, each of the five stimulus elements (1-5) were mapped onto one of four vocabularies, corresponding to four types of modality / format. Importantly, the timing of all stimuli was equated across these four conditions in order to make comparisons across conditions viable.

Auditory. For the auditory conditions, the stimulus elements consisted of pure tones of various frequencies (1 = 261.6 Hz, 2 = 277.2 Hz, 3 = 349.2 Hz, 4 = 370 Hz, and 5 = 493.9 Hz) corresponding to musical notes C, C#, F, F#, and B, respectively, played through headphones. For the “slow” and “fast” input conditions, each stimulus element (tone) had durations of 250 ms and 125 ms, respectively. As an example, the sequence 1-2-1-3 consists of the following four notes in this order: C, C#, C, and F.

Table 3.2

Test Items

<u>Legal</u>	IAS	FAS	<u>Illegal</u>	IAS	FAS
4-1-3-5-2-3	2.5	2.5	1-4-5-1-3-3	0	0
1-2-1-3-5-2	4.5	2.0	4-5-1-2-1-3	0	2.0
4-3-5-2-5-5	2.0	1.5	4-2-1-3-1-5	0	0
4-1-3-5-2-5	2.5	2.0	1-5-3-3-2-2	0	0
4-1-1-1-3	2.0	2.0	1-5-3-4-2	0	0
1-2-1-1-3	4.5	2.0	4-2-1-5-3	0	0
1-2-3-5-2	5.0	2.0	1-5-3-1-2	0	0
4-1-1-3	2.0	2.0	4-5-1-3	0	1.5
4-3-5-2	2.0	2.0	4-5-2-2	0	0
1-2-5	4.5	1.0	1-4-2	0	0

Note: The numbers refer to a particular visual or auditory stimulus (see text). IAS = initial anchor strength; FAS = final anchor strength.

Visual-temporal. For the visual-temporal conditions, the stimulus elements consisted of different colored squares (1=red, 2=blue, 3=yellow, 4=green, 5=black) appearing sequentially in the center of the computer screen at approximately eye level. Each square (2.6 cm x 2.6 cm) appeared for 250 ms in the “slow” input condition and 125 ms in the “fast” condition. Thus, the sequence 1-2-1-3 consists of a temporal sequence of red, blue, red, and yellow squares (see Figure 3.2).

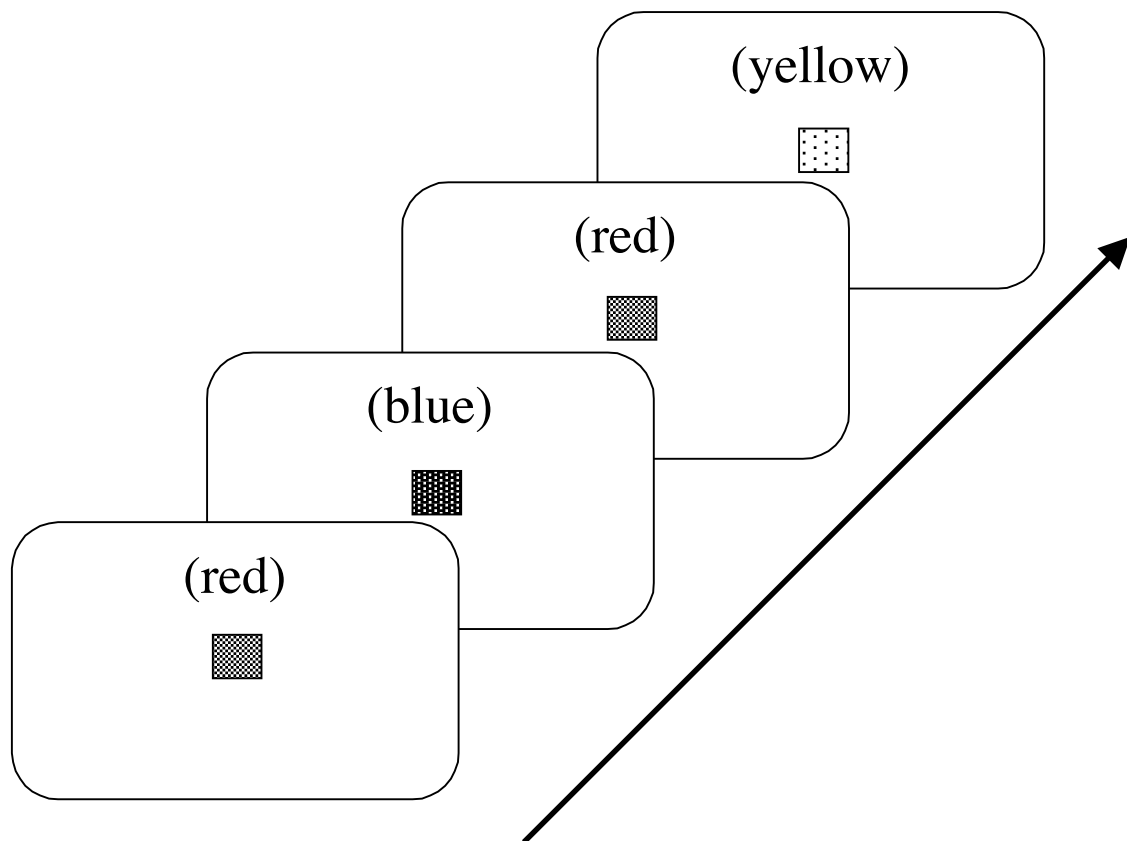


Figure 3.2. Example of visual-temporal sequence 1-2-1-3. The arrow designates the flow of time. Note that participants did not see the labels for each color; they are included here only for illustrative purposes.

Visual-spatial. For the visual-spatial conditions, the stimulus elements consisted of the same colored squares described above, except that all squares in a sequence were presented simultaneously along a horizontal row, from left to right, at approximately eye level. The timing of the stimuli was equal to the cumulative presentation time of the visual-temporal stimuli. That is, in the slow condition, a visual-spatial sequence was displayed for a number of msec equal to $250 \times N$, where N is the number of squares in the sequence. For the fast condition, a sequence was displayed for $125 \times N$ ms. Thus, the visual-spatial sequence 1-2-1-3 consists of a row of squares appearing simultaneously

for $(250 \times 4) = 1000$ ms or $(125 \times 4) = 500$ ms, from left to right: red, blue, red, yellow (see Figure 3.3).

Visual-spatiotemporal. The visual-spatiotemporal conditions were designed to include both temporal and spatial information in each sequence. Each stimulus element consisted of a black square, displayed on the computer monitor in different locations (element 1 represents the leftmost location, and 5 the rightmost). Each black square (2.6×2.6 cm) was positioned in a horizontal row across the middle of the screen at approximately eye level, with 2.5 cm separating each position. Each element appeared for 250 ms or 125 ms, depending on the presentation rate condition. Figure 3.4 illustrates the visual-spatiotemporal sequence 1-2-1-3.

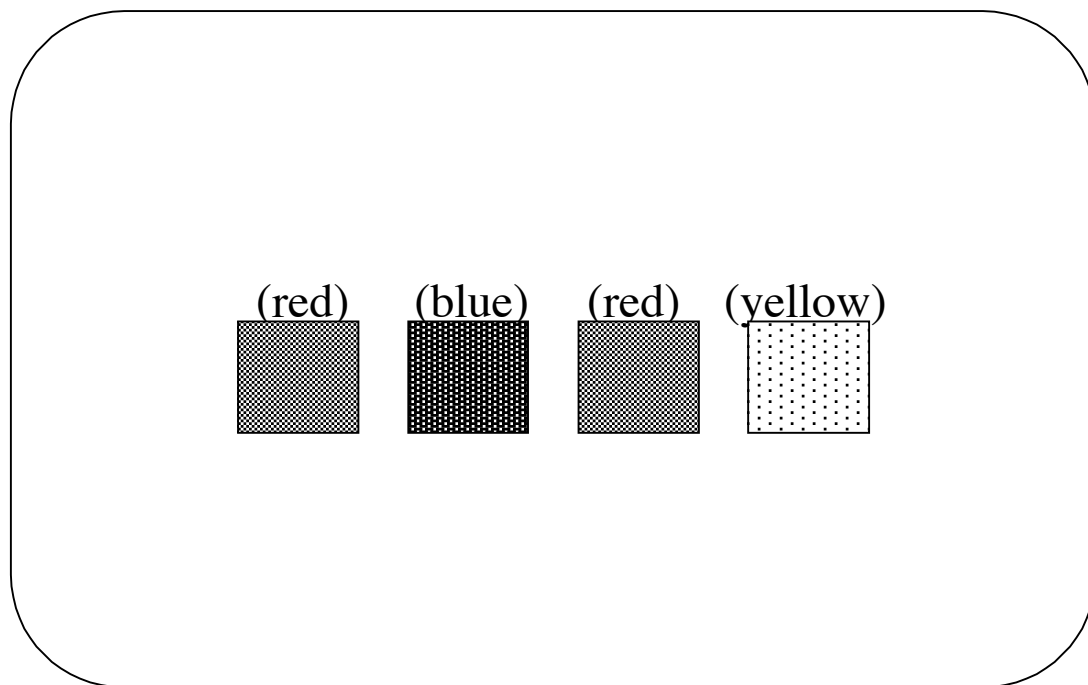


Figure 3.3. Example of visual-spatial sequence 1-2-1-3. Note that participants did not see the labels for each color; they are included here only for illustrative purposes.

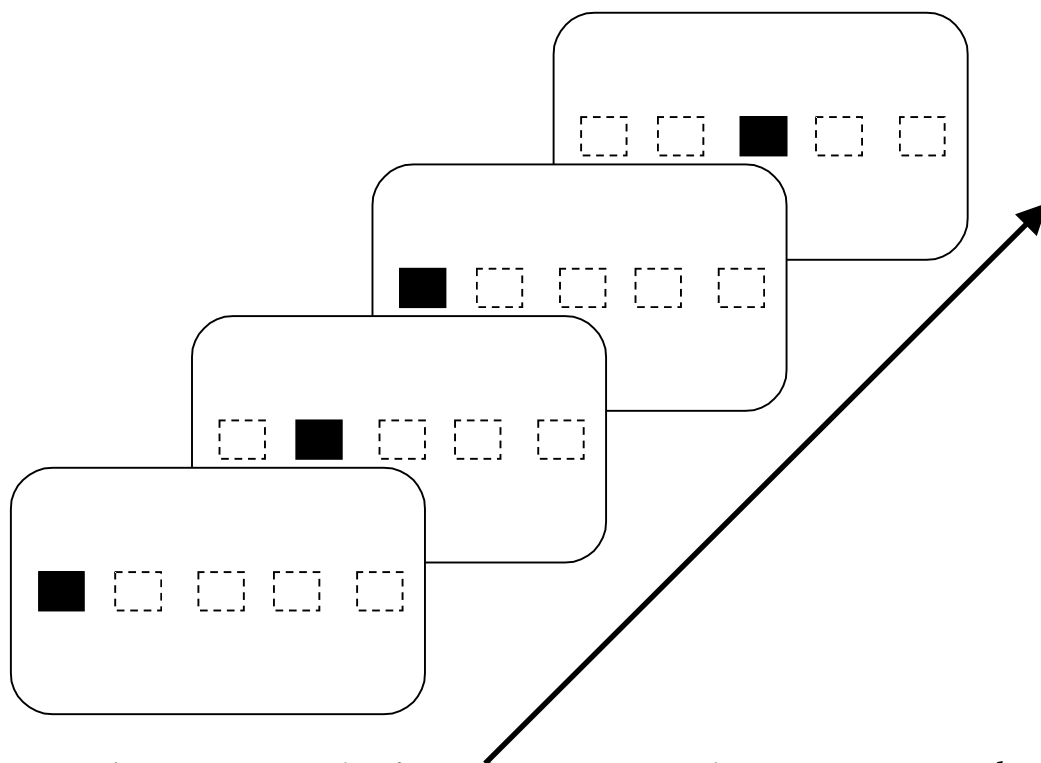


Figure 3.4. Example of visual-spatiotemporal sequence 1-2-1-3. The arrow designates the flow of time. Note that participants did not see the dashed squares, which are included here only for illustrative purposes.

Procedure

Participants were assigned randomly to one of twelve conditions, four control groups and eight experimental groups. Within the experimental groups, the 4 x 2 design consisted of two factors: Modality/Format (auditory, visual-temporal, visual-spatial, and visual-spatiotemporal) and Presentation Rate (slow and fast). The slow and fast groups received the input material at a presentation rate of 4 and 8 elements per second, respectively (corresponding to individual element durations of 250 ms and 125 ms).

The four control groups consisted of the four different Modality/Format conditions; all control groups received the input at the slow

presentation rate only.¹⁰ The control groups were meant to ensure that any learning that the experimental groups displayed was dependent upon the exposure to the input material in the training phase. Consequently, the experimental groups received both training and test phases whereas the control groups received the test phase only.

At the beginning of the training phase, the slow and fast experimental group participants were instructed that they would hear or see pairs of sequences. For each pair of sequences (listed in Table 3.1), they had to decide whether the two sequences were the same or not and indicate their decision by pressing a button marked *YES* or *NO*. This match–mismatch paradigm, also used by Conway and Christiansen (2005), served as a way to encourage participants to pay attention to the stimuli without giving them explicit instruction that the sequences conformed to an underlying structure. Note that there is no feedback given during the task.

Each pair was presented six times in random order for a total of 72 exposures, with the timing parameters described earlier. In all conditions, a 2-s pause occurred between the two sequences of each pair and following the last sequence of the pair. After exposure to the two sequences, a prompt was displayed on the computer monitor asking for the participant's response, and it stayed on the screen until a button press was made. After another 2-s pause, the next training pair was presented. The entire training phase in each condition lasted roughly 10 min for each participant.

Before the beginning of the test phase, the experimental group participants were told that the sequences they had just observed had been

¹⁰ We only incorporated the slow presentation rate for the control groups because, assuming that the control groups would perform at chance levels, there is no reason to believe that increasing the rate of presentation would change performance levels.

generated by a computer program that determined the order of the stimuli by using a complex set of rules. They were told that they would now be presented with new sequences. Some of these would be generated by the same program, whereas others would not be. It was the participant's task to observe each of the new sequences and then classify it without feedback in terms of whether it was generated by the same rules by pressing a button marked either *YES* or *NO*. The control participants, who did not participate in the training phase, received an identical test task.

The 20 test sequences were presented one at a time, in random order, to each participant. The timing of the test sequences was the same as that used during the training phase (250-ms or 125-ms pulse durations and 2-s pauses before and after each sequence).

Results

The training performances for the eight experimental groups are shown in Table 3.3, which reports the mean number of correct match/mismatch decisions out of 72. We submitted the data to a two-way ANOVA with the factors Modality/Format and Presentation Rate. There was a main effect of Modality/Format, $F(3,88) = 26.1$, $p < .0001$ but no main effect of Presentation Rate, $F(1,88) = .28$, $p = .60$, nor a significant interaction, $F(3,88) = .55$, $p = .55$. Recall that the training task was merely meant to encourage participants' attention to the training stimuli and thus comparisons of performance between different conditions is of minimal importance for our purposes. The data of more interest are the test classification results, which provide a measure of the extent to which participants can extract statistical regularities and generalize this knowledge to novel patterns.

Table 3.3

Experimental Group Training Results

Modality/Format	Slow (4 elements/s)		Fast (8 elements/s)	
	<i>M</i>	%	<i>M</i>	%
Auditory	49.4	68.7	50.4	70.0
Visual-Temporal	44.1	61.3	42.8	59.4
Visual-Spatial	47.3	65.8	48.7	67.6
Visual-Spatiotemporal	40.8	56.6	41.5	57.7

Note. Mean values reported out of 72 possible correct.

The control group test results are shown in Table 3.4, which reports the mean number of correct classification responses out of 20. As expected, none of the four groups performed greater than chance levels, thus indicating that any learning displayed by the experimental groups is due to exposure to the material during the training phase.¹¹

¹¹ The visual-spatiotemporal control group actually performed statistically worse than chance, indicating that the visual-spatiotemporal experimental groups might be biased against an effect of learning.

Table 3.4

Control Group Test Results

<u>Modality/Format</u>	<u>M</u>	<u>%</u>	<u>t(11)</u>
Auditory	10.8	54.2	1.76
Visual-Temporal	9.9	49.6	0.16
Visual-Spatial	9.3	46.7	-0.83
Visual-Spatiotemporal	8.1	40.4	-2.73*

Note. Mean values reported out of 20 possible correct. T-tests are conducted with respect to chance levels.

* $p < .05$

The experimental group test results are shown in Table 3.5 and displayed graphically in Figure 3.5. Notice first that the auditory, visual-temporal, and visual-spatial groups all evidenced learning above chance levels in the slow presentation conditions. Notice, too that only the visual-temporal group's performance declined at the fast presentation rate, dropping to chance levels.

Table 3.5

Experimental Group Test Results

Modality/Format	Slow (4 elements/s)			Fast (8 elements/s)		
	<i>M</i>	%	<i>t</i> (11)	<i>M</i>	%	<i>t</i> (11)
Auditory	13.9	69.5	5.42***	14.3	71.7	6.28***
Visual-Temporal	12.2	60.8	3.17**	10.4	52.1	0.92
Visual-Spatial	13.1	65.4	3.39**	12.6	62.9	3.11**
Visual-Spatiotemporal	10.9	54.6	1.35	11.5	57.5	3.00*

Note. Mean values reported out of 20 possible correct. T-tests are conducted with respect to chance levels.

* $p < .05$. ** $p < .01$. *** $p < .001$

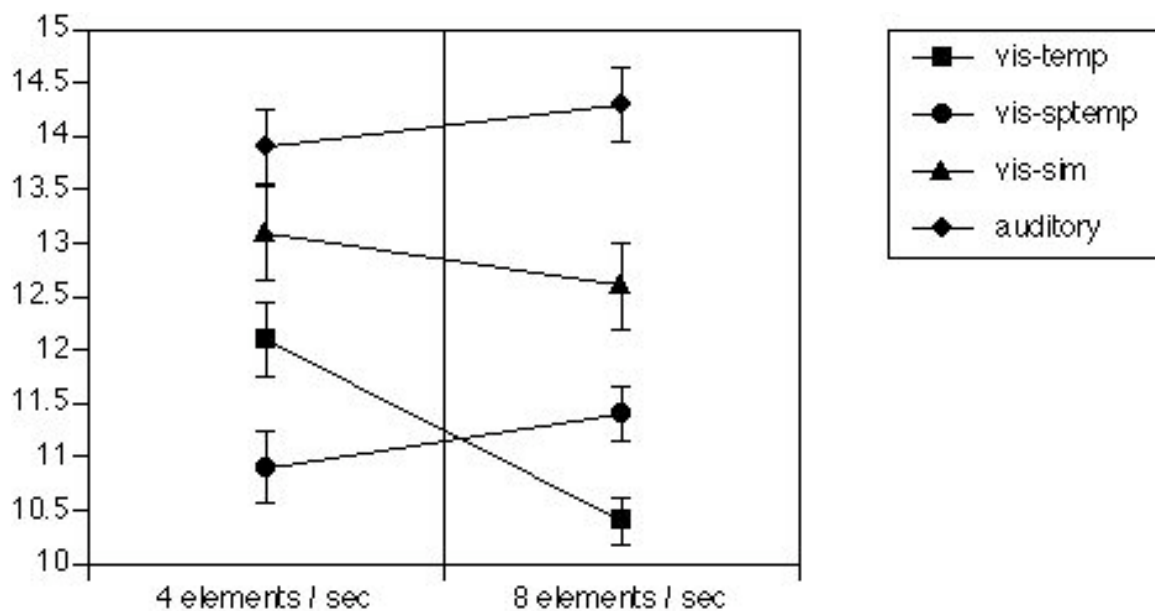


Figure 3.5. Experimental group test performances, showing number of correct test items out of 20, and including error bars.

We submitted the experimental group data to a two-way ANOVA with the factors Modality / Format and Presentation Rate. There was a main effect of Modality / Format, $F(3,88) = 7.96$; $p < .0001$, but no main effect of Presentation Rate, $F(1,88) = 0.40$; $p = .52$, nor a significant interaction, $F(3,88) = 1.18$; $p = .32$.

We continued with planned comparisons between the four different modality / formats, collapsed across presentation rate. The results revealed that auditory performance was significantly greater than both visual-temporal ($p < .0001$) and visual-spatiotemporal performance ($p < .001$), as well as marginally greater than visual-spatial performance ($p = .067$). Visual-spatial performance in turn was significantly greater than both visual-temporal ($p < .05$) and visual-spatiotemporal performance ($p < .05$). There were no differences between the visual-temporal and visual-spatiotemporal groups ($p = .91$). We also conducted planned comparisons within each modality / format between the slow and fast conditions. Confirming our earlier observation, only the visual-temporal group showed a difference in test performance, scoring significantly lower at the fast presentation rate ($p < .05$).

These results suggest *quantitative* learning differences between sense modalities. In order to detect more subtle modality-specific learning differences, we conducted item analyses to determine whether particular test items were more or less difficult to classify in each modality / format. We conducted two separate two-way ANOVA's (Modality / Format x Test Item), one for each rate of presentation, entering the test performance data averaged across subjects for each test item. For the slow presentation rate, there were main effects both of Test Item, $F(19, 880) = 3.47$, $p < .0001$, and Modality / Format, $F(3, 880) = 4.53$, $p < .01$, but no interaction, $F(57, 880) = 1.17$, $p = .19$. For the fast presentation rate, there also were main effects for Test

Item, $F(19, 880) = 3.39, p < .0001$, and Modality / Format, $F(3, 880) = 8.18, p < .0001$, as well as a significant interaction, $F(57, 880) = 2.14, p = .001$.

The significant interaction in the item analysis for the fast presentation conditions suggests that there exist learning-related *qualitative* modality effects. We followed up with regression analyses to determine which sources of information may have been extracted by each modality / format condition. Each of the legal and illegal test items were assessed in terms of their initial and final anchor strengths (IAS and FAS), an indication of the relative frequencies of the initial and final fragment “chunks” (i.e., bi- and tri-grams) that exist in similar positions in the training items (Conway & Christiansen, 2005). For example, the test item 1-2-1-3-5-2 has an IAS of 4.5 and an FAS of 2.0, indicating that the initial chunks 1-2, 2-1, and 1-2-1 occur frequently in the initial positions of the training set, whereas the final chunks 3-5, 5-2, and 3-5-2 occur slightly less frequently in the final positions of the training set.¹² We used IAS, FAS, and item length as predictors in four different multiple regression analyses, one for each modality / format condition (each collapsed across presentation rate). The results of the multiple regression analyses should indicate which of these three measures best predicts whether a participant in each modality / format condition will endorse a test item as legal. The results reveal a striking difference between the auditory and the visual conditions: FAS was a significant predictor for auditory endorsements ($p < .005$), whereas IAS was a significant and marginally significant predictor

¹² In Conway and Christiansen (2005), we assessed these stimuli in terms of additional information sources, such as novelty (Redington & Chater, 1996), novel fragment position (Johnstone & Shanks, 1999), and similarity (Vokey & Brooks, 1992), but found that for this particular test and training set, these measures were highly correlated with IAS and FAS. A principle components analysis revealed that all these sources could be efficiently reduced to IAS, FAS, and item length. Therefore, we include only those three measures in the regression analyses.

for visual-spatial ($p < .05$) and visual-temporal endorsements ($p = .09$), respectively. Length was a significant and marginally significant predictor for visual-spatial ($p < .05$) and visual-temporal endorsements ($p = .087$), respectively. It is important to note that there are no length differences between legal and illegal items. Thus, participants in the visual-spatial and visual-temporal conditions apparently were partially relying on an incorrect strategy to guide their test classifications.

In summary, for the experimental group test data, there was a main effect of modality / format, with auditory performance greatest, closely followed by visual-spatial performance, and with the visual-temporal and visual-spatiotemporal groups performing worst. We also found that at the fast presentation rate, only visual-temporal performance suffered. These quantitative learning differences were accompanied by qualitative learning effects. Consistent with previous results (Conway & Christiansen, 2005), audition and vision were differentially biased toward the end and beginning of input sequences, respectively.

These results provide key support for a modality-constrained view of statistical learning. Participants in the visual conditions appeared to best extract statistical patterns when the input was presented in a spatial format rather than a temporal or spatiotemporal one. Additionally, visual learning relied upon statistical information present at the beginning of input sequences. In contrast, auditory learning excelled in the encoding of temporal input, mediated by greater sensitivity to the statistical structure of the end of input sequences. Furthermore, the quantitative modality constraints were magnified at the highest presentation rate, with visual-temporal learning dropping in

overall performance, highlighting the fact that vision is poor at encoding temporal regularities, especially at fast presentation rates.

General Discussion

The modality constraints observed here extend the results of Conway and Christiansen (2005), who found that auditory statistical learning of sequential structure exceeded that of visual or tactile learning, as well as that of Saffran (2002), who found that learning of predictive relationships occurred better for visual-simultaneous and auditory input compared to visual-sequential material. In the current study, visual learning nearly matched auditory learning but only when the visual regularities were presented in a simultaneous/spatial format. These results illustrate that the extraction of statistical patterns is heavily affected by the modality and presentation format in which the input is delivered. Furthermore, we found that increasing the rate of presentation magnified these modality constraints, causing a decline in learning performance for the visual-temporal condition but not for the other formats.

These data suggest that a full understanding of statistical learning, and likely other forms of perception, learning, and memory, will only be borne out by taking into account effects of modality in relation to the dimensions of space and time. The empirical evidence supports what we call a *modality-constrained view* of perception and cognition. That is, consistent with previous work in perception (Mahar et al., 1994), learning (Saffran, 2002), and short-term memory (Penney, 1989), we suggest that performance on particular tasks will vary depending on the sensory modality and the input format (i.e., spatial, temporal, or spatiotemporal) in question.

Before attempting to further clarify the nature of this modality-constrained view, it is worthwhile to consider possible objections. Although the modality effects reviewed here and elsewhere appear to be substantial and robust, there also exist commonalities between visual and auditory processing (e.g., Kubovy & Van Valkenburg, 2001; Marks, 1978). Both vision and audition can localize stimuli in space, can detect movement, can perceive rhythms and sequential patterns, and can discriminate objects based on when they occurred. As the Gibsons (1966; 1969) argued, such environmental information might be “amodal” in the sense that these and other features can be picked up across multiple sense modalities. If so, then different stimulus energies could be considered to be equivalent, or invariant, to one another. Similarly, Marks (1978) argued that the sense modalities have much in common in terms of phenomenological attributes, principles, and mechanisms. As he suggested, it is likely that the auditory and visual systems rely upon many of the same computational algorithms and neural architectures (see Shamma, 2001). From these perspectives, perhaps the vision-space, audition-time analogy is misleading (Handel, 1988).

We agree that it may be misleading to consider vision solely within a spatial framework and audition as existing only within a temporal one. However, there appear to be real biases in terms of which dimensions are more or less important for each modality. From a purely phenomenological perspective, it seems nearly impossible to imagine an a-temporal sound or a non-spatial visual percept. On the other hand, it is easier to think of sound without space or a visual scene without the passage of time. Time appears to be the primary gridwork for audition, with sounds changing in certain ways

over time, whereas space is the primary referent for vision, with visual objects defined by size and shape (Hirsh, 1967).

More importantly, the empirical data shown here and elsewhere (e.g., Friedes, 1974) act as a counter-argument to the possibility of amodal cognitive processes. As Barsalou (1999) argued, modality differences make it difficult to cling to amodal theories of perception and cognition. Though it may be true that both vision and audition can perceive the invariant structure of certain features such as temporal patterns or object locations, nevertheless, each sense appears to be differentially adept at detecting a subset of stimulus characteristics, including temporal and spatial statistical patterns.

Thus, however much the sensory systems may be similar, it is clear that certain constraints affect their processing, and therefore it is important to understand what causes the constraints. Within the context of verbal short-term memory, Penney (1989) suggested that auditory and visual stimuli are processed in separate “streams” with each having different properties and capabilities. The auditory stream is characterized by strong associations between successive/sequential items, whereas the visual stream encodes associations between simultaneously presented spatial input. This makes sense in light of what is known about the principles of auditory grouping: sounds that rapidly follow one another tend to be produced by the same environmental source and thus are perceived or “grouped” together, whereas it is spatial rather than temporal contiguity which is important for visual perception (Goldstein, 2002). Possibly, a similar associative neural learning mechanism may be operative in both the auditory and visual cortical areas, but each makes its computations over a different set of input. The existence of similar underlying computational principles helps explain the processing

similarities across the modalities, while the proposed separate processing streams allow for the existence of modality-specific differences. Under this framework, visual mechanisms might associate features and/or objects that are close together in terms of their location along the retinotopic map (i.e., in terms of space). Auditory mechanisms would encode relations among events occurring close together in time.

This modality-constrained framework helps explain the current data. First, the data revealed that subjects were best able to encode statistical regularities contained within visual input when the material was presented spatially, rather than spatiotemporally or temporally. In contrast, subjects excelled at the processing of auditory input (presented temporally), outperforming all other visual conditions. Second, when the input stream was presented at relatively fast rates (8 elements per second), modality effects were magnified: subjects in the visual-temporal condition suffered a substantial decrement in performance.

The differences in the initial/final sensitivities by vision and audition were also observed previously (Conway & Christiansen, 2005). It is important to note that for the current stimulus set, both fragment-initial and fragment-final information are equally helpful for making correct classification judgments; thus, it isn't likely that the observed qualitative sensitivity biases are due to experiment-specific learning. Instead, it is more likely that learners come into the experiment with pre-existing auditory-final and visual-initial encoding biases. Interestingly, similar modality initial/final effects are seen in the realm of serial recall (i.e., serial position effects) with audition showing better recall for the end of lists (Crowder, 1986) and under certain circumstances, vision showing better recall for the beginning of lists (Beaman,

2002). Thus, there may exist a global perceptual / cognitive constraint on vision and audition, evolved through currently unknown selection pressures, to differentially attend to information in initial and final sequence positions, respectively.

As others have argued (Barsalou, 1999; Glenberg, 1997; Penney, 1989; Tulving & Madigan, 1970), understanding the role played by the sensory systems is key to having a full theory of human memory and perception, and any theory of cognition will not be adequate if it cannot account for modality-specific phenomena. Our data shed light on one important modality constraint: implicit statistical learning is heavily constrained by the sense modality and presentation format in which the patterns occur. These constraints suggest that auditory and visual statistical learning occurs in separate, modality-specific streams, each having different biases and properties. Additional research must further elucidate the nature of each modality-constrained learning system and how they support human cognition more broadly.

References

- Altmann, G.T. (2002). Statistical learning in infants. *Proceedings of the National Academy of Sciences*, 99, 15250-15251.
- Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences*, 22, 577-609.
- Beaman, C. P. (2002). Inverting the modality effect in serial recall. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 55A, 371-389.
- Bushara, K.O., Weeks, R.A., Ishii, K., Catalan, M.-J., Tian, B., Rauschecker, J.P., & Hallett, M. (1999). Modality-specific frontal and parietal areas for auditory and visual spatial location in humans. *Nature Neuroscience*, 2, 759-766.
- Collier, G.L. & Logan, G. (2000). Modality differences in short-term memory for rhythms. *Memory & Cognition*, 28, 529-538.
- 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.
- Crowder, R. G. (1986). Auditory and temporal factors in the modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2, 268-278.
- Fiser, J. & Aslin, R.A. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12, 499-504.

- Freides, D. (1974). Human information processing and sensory modality: Cross-modal functions, information complexity, memory, and deficit. *Psychological Bulletin*, 81, 284-310.
- Geldard, F.A. (1970). Vision, audition, and beyond. In W.D. Neff (Ed.), *Contributions to sensory physiology, Vol.4* (pp. 1-17). New York: Academic Press.
- Gibson, E.J. (1969). *Principles of perceptual learning and development*. Englewood Cliffs, NJ: Prentice-Hall.
- Gibson, J.J. (1966). *The senses considered as perceptual systems*. Boston: Houghton-Mifflin.
- Glenberg, A.M. (1997). What memory is for. *Behavioral and Brain Sciences*, 20, 1-55.
- Glenberg, A.M. & Swanson, N.C. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 12, 3-15.
- Goldstein, E.B. (2002). *Sensation and perception* (6th edition). USA: Wadsworth.
- Gomez, R.L. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology*, 33, 154-207.
- Gomez, R.L. & Gerken, L.A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70, 109-135.
- Handel, S. (1988). Space is to time as vision is to audition: Seductive but misleading. *Journal of Experimental Psychology: Human Perception & Performance*, 14, 315-317.
- Handel, S., & Buffardi, L. (1969). Using several modalities to perceive one temporal pattern. *Quarterly Journal of Experimental Psychology*, 21, 256-266.
- Hirsh, I.J. (1967). Information processing in input channels for speech and

- language: The significance of serial order of stimuli. In F.L. Darley (Ed.), *Brain mechanisms underlying speech and language* (pp. 21-38). New York: Grune & Stratton.
- Johnstone, T., & Shanks, D. R. (1999). Two mechanisms in implicit artificial grammar learning? Comment on Meulemans and Van der Linden (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 524-531.
- Kubovy, M. (1988). Should we resist the seductiveness of the space:time::vision:audition analogy? *Journal of Experimental Psychology: Human Perception and Performance*, 14, 318-320.
- Kubovy, M. & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, 80, 97-126.
- Mahar, D. Mackenzie, B. & McNicol, D. (1994). Modality-specific differences in the processing of spatially, temporally, and spatiotemporally distributed information. *Perception*, 23, 1369-1386.
- Marks, L.E. (1978). *The unity of the senses: Interrelations among the modalities*. New York: Academic Press.
- O'Connor, N. & Hermelin, B. (1978). *Seeing and hearing and space and time*. New York: Academic Press.
- Penney, C.G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, 17, 398-422.
- Potter, M. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 509-522.
- Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Behavior*, 6, 855-863.
- Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A

- reevaluation. *Journal of Experimental Psychology: General*, 125, 123–138.
- Repp, B.H. & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 1085-1099.
- Saffran, J.R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, 47, 172-196.
- 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.
- Seifritz, E., Esposito, F., Neuhoff, J.G., & Di Salle, F. (2003). Response: Sound analysis in auditory cortex – from temporal decomposition to perception. *Trends in Neurosciences*, 26, 231-232.
- Shamma, S. (2001). On the role of space and time in auditory processing. *Trends in Cognitive Sciences*, 5, 340-348.
- Stoffregen, T.A. & Bardy, B.G. (2001). On specification and the senses. *Behavioral & Brain Sciences*, 24, 195-261.
- Tulving, E. & Madigan, S.A. (1970). Memory and verbal learning. *Annual Review of Psychology*, 21, 437-484.
- Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 328–344.

CHAPTER FOUR

Statistical Learning Within and Between Modalities: Pitting Abstract Against Stimulus-Specific Representations

A core debate in the psychological sciences is to what extent acquired knowledge consists of modality-dependent versus abstract or amodal representations. Traditional information processing approaches to cognition have emphasized the operation of amodal symbol systems (Fodor, 1975; Pylyshyn, 1984), whereas more recently, the “embodiment” approach proposes instead that cognition is grounded in modality-specific sensorimotor mechanisms (Barsalou, 1999; Barsalou, Simmons, Barbey & Wilson, 2003; Glenberg, 1997; Lakoff 1988). This debate has especially come to a head in the area of implicit statistical learning of artificial grammars.¹³ In Arthur Reber’s (1967; 1969) early work, letter strings were generated from an artificial grammar. Because the strings obeyed the overall rule structure of the grammar, they were constrained in terms of what letters could follow other letters. Participants not only showed evidence of learning this structure “implicitly”, but they could apparently transfer their knowledge of the legal regularities from one letter vocabulary (e.g., M, R, T, V, X) to another (e.g., N, P, S, W, Z) as long as the same underlying grammar was used for both (Reber, 1969). This effect has been replicated many times, with transfer being demonstrated not just across letter sets (e.g., Brooks & Vokey, 1991; Mathews

¹³ Artificial grammar learning is “statistical” in the sense that successful test performance can be achieved by encoding something akin to chunks of elements (Perruchet & Pacteau, 1990) or learning the transitional probabilities among consecutive elements (Saffran, Johnson, Aslin, & Newport, 1999).

et al., 1989; Shanks, Banbury, & Henry, 1997) but also across sense modalities (Altmann, Dienes, & Goode, 1995; Manza & Reber, 1997; Tunney & Altmann, 2001). Both adults and infants have displayed successful transfer (Gomez & Gerken, 1999; Marcus, Vijayan, Rao, & Vishton, 1999).

The usual way to explain transfer effects in artificial grammar learning (AGL) tasks is by proposing that the learning is based on “abstract” knowledge, that is, knowledge not directly tied to the surface features or sensory instantiation of the stimuli (Altmann et al., 1995; Marcus et al., 1999; Pena, Bonatti, Nespor, & Mehler, 2002; Reber, 1989; 1993; Shanks et al., 1997). For instance, the human cognitive system might encode patterns among stimuli in terms of “abstract algebraic-like rules” that encode relationships between amodal variables (Marcus et al., 1999, p. 79). Such a proposal emphasizes the structural relationships among items and deemphasizes information pertaining to specific features of the stimulus elements (see Reber, 1993, pp. 120-121). Alternatively, instead of abstract knowledge, participants may be learning the statistical structure of the input sequences in a modality- or stimulus-specific manner (e.g., Chang & Knowlton, 2004; Christiansen & Curtin, 1999; Conway & Christiansen, 2005; McClelland & Plaut, 1999; Perruchet, Tyler, Galland, & Peereman, 2004; Seidenberg & Elman, 1999).

In this chapter, we present new evidence from a set of AGL experiments supporting a modality-constrained or embodied view of statistical learning. 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 dual-grammar cross-over design such that the grammatical test sequences of one grammar were used as the ungrammatical test sequences for the other grammar. For example, in Experiment 4,

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. Crucially, for each participant, all test items from both grammars were instantiated only visually or only aurally. This cross-over design allows us to make the following prediction. If participants learn the abstract rules underlying both grammars, they ought to classify all sequences as equally grammatical (by our scoring scheme, achieving 50% performance). 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, leading to above-chance performance. To preview, the data from these experiments suggest that learners encoded the sequential patterns and generalized their knowledge to novel instances by relying on stimulus-specific, not abstract, representations.

Experiment 4: 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. In our scoring scheme, a null effect of learning (50% performance) can mean one of two possibilities: 1) participants were unable to display adequate knowledge of the statistical regularities; or 2) participants could learn the regularities but

the knowledge exists in an amodal format that does not retain information regarding the sense modality of the input. Accordingly, if performance levels are significantly above chance, it shows that participants learned the statistical regularities from both grammars in a modality-specific manner. In order to compare dual-grammar learning to “standard” AGL learning performance, we employed single-grammar, unimodal learning conditions as a baseline.

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 4.1), were used to generate two sets of non-overlapping stimuli (see Table 4.1). 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. For each participant, one grammar was instantiated as a color vocabulary, five differently colored squares, and the other grammar was instantiated as an auditory vocabulary, five pure 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.

All visual stimuli were presented in a serial 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 500 ms with 100 ms occurring between elements. Each sequence was separated by 1700 ms of blank screen.

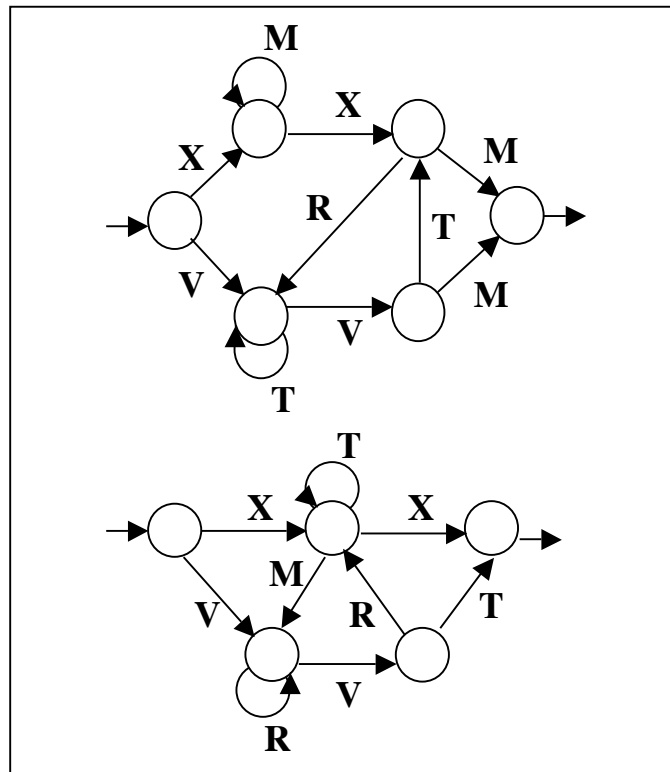


Figure 4.1. Grammar A (top) and Grammar B (bottom) used in all three experiments. The letters from each grammar were instantiated as colors or tones (Experiment 4), colors or shapes (Experiment 5A), tones or nonwords (Experiment 5B), two different shape sets (Experiment 6A), or two different nonword sets (Experiment 6B).

Table 4.1

Training and Test Items for Grammars A and B

	Grammar A	Grammar B
Training	VVM	XXM
	XXM	VTRM
	XXRVM	VVRXM
	VTVTM	XTTXM
	XXRVTM	XMRTRM
	VTTVTM	VRRTRM
	XMMMXXM	XMRRTRM
	XXRTTVM	VRVRTXM
	VTVTRVM	VRRVRXM
Test	VTVM	XTXM
	VVTM	VRTRM
	XMMXXM	XTMTRM
	VTTVM	VRVRXM
	VTTTVM	VVRTXM
	VVTRVM	XMVRXM
	VTTTVTM	VRRRTRM
	XXRTVTM	XMVRTXM
	XXRTTVM	VVRTTXM
	XMMXRVM	XMRVRXM

Procedure. Participants were randomly assigned to one of four conditions, two experimental and two baseline. Participants in the experimental conditions were exposed to the training sequences from both grammars, with one grammar instantiated as the color vocabulary and the other as tones. Modality-grammar assignments were counter-balanced across participants. Additionally, the particular assignment of letters to visual or auditory elements within each grammar was randomly determined for each participant.

At the beginning of the experiment, participants were told that they would hear and/or see sequences of auditory and visual stimuli. They were told that it was important to pay attention to the stimuli because afterward they would be tested on what they observed. Importantly, the participants were not explicitly told of the existence of the grammars, underlying rules, or regularities of any kind. 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. Figure 4.2 shows a schematic of a possible sample of the stimulus presentation.

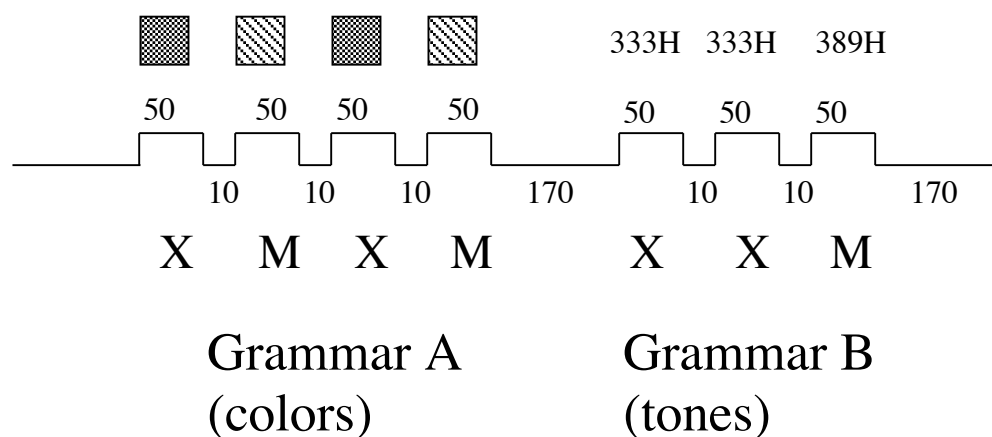


Figure 4.2. Sample stimulus presentation for Experiment 4. Sequences from the two grammars were interleaved randomly. For each participant, one grammar was instantiated with the color vocabulary and the other grammar instantiated with the tone vocabulary. The letters denote a particular color or tone, depending on the grammar and vocabulary. The timing of the sequence elements are shown, in milliseconds.

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 a new set of color or tone sequences. 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). For scoring purposes, test sequences from a particular grammar that were instantiated in the same sense modality as in the training phase were deemed “grammatical” whereas the sequences from the other grammar were deemed “ungrammatical”. Thus, 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 single-grammar 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, counter-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 as the experimental participants, instantiated with the same vocabulary with which they were trained on. Thus, the baseline conditions (Visual-Baseline and Auditory-Baseline conditions) assess visual and auditory learning with one grammar alone, in a way similar to the standard AGL design.

Results and Discussion

Table 4.2 reports mean correct classification scores (out of 20) for each group as well as t-tests compared to chance levels. Each group’s overall performance was better than 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, $t(9) = 1.1$; $p = .30$.

Table 4.2

Experiment 4 Mean Performance and Tests of Significance versus Chance

Modality	<u>Experimental (Dual-grammar)</u>			<u>Baseline (Single-grammar)</u>		
	<i>M</i>	%	<i>t</i> (9)	<i>M</i>	%	<i>t</i> (9)
Visual	12.7	63.5	2.76*	12.4	62.0	2.54*
Auditory	14.1	70.5	4.38**	13.1	65.5	3.44**

Note. Mean values reported out of 20 possible correct. T-tests are conducted with respect to chance levels.

* $p < .05$. ** $p < .01$

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. Perhaps surprisingly, the levels of performance in the dual-grammar experimental conditions were 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 (according to our scoring scheme). 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 would be as perceptually dissimilar as possible, using colors and shapes (Experiment 5A) and tones and nonwords (Experiment 5B).

Experiment 5: Intramodal Learning Along Different Perceptual Dimensions

The purpose of Experiment 5 was to further explore the stimulus-specific nature of statistical learning. Specifically, we assessed whether participants can learn two sets of statistical regularities when they are presented within the same sense modality but instantiated along two different perceptual “dimensions.” Experiment 5A examines intramodal learning in the visual modality while Experiment 5B examines auditory learning. For Experiment 5A, one grammar was instantiated with colors and the other with shapes. For Experiment 5B, one grammar was instantiated with tones and the other with nonwords.

Method

Subjects. For Experiment 5, 60 additional participants (10 in each condition) were recruited in the same manner as in Experiment 4.

Materials. Experiment 5 incorporated the same two grammars, training and test sequences that were used in Experiment 4. Experiment 5A used two visual vocabularies. The first was the same set of colors used in Experiment 4;

the second consisted of five abstract, geometric shapes. These shapes were chosen to be perceptually distinct yet not amenable to a verbal coding strategy. Experiment 5B used two auditory vocabularies. The first consisted of the same set of tones used in Experiment 4; the second consisted of five different nonwords, recorded as individual sound files spoken by a human speaker (from Gomez, 2002): “vot,” “pel,” “dak,” “jic,” and “rud.”

Procedure. Participants were randomly assigned to one of six conditions, two for Experiment 5A, two for Experiment 5B, and two new single-grammar baseline conditions. The general procedure was the same as in Experiment 1. In Experiment 5A, participants were trained on the two visual grammars and then tested on their ability to classify novel sequences. For half of the participants, these test sequences were instantiated using the color vocabulary while for the other half they were instantiated using the shape vocabulary. Likewise, in Experiment 5B, participants were trained on both auditory grammars and then at test, half of the participants received all test sequences as tones and the other half received 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 same color and tone vocabulary baseline data from Experiment 4). In all other respects, the procedure for Experiment 5 was the same as in Experiment 4.

Results and Discussion

Mean scores and t-tests compared to chance levels are shown in Table 4.3. Each group’s overall performance was better than 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, $t(9) = -.178, p = .86$.

Table 4.3

Experiment 5 Mean Performance and Tests of Significance versus Chance

Dimension	<u>Experimental (Dual-grammar)</u>			<u>Baseline (Single-grammar)</u>		
	<i>M</i>	%	<i>t</i> (9)	<i>M</i>	%	<i>t</i> (9)
Experiment 5A						
Colors	11.9	59.5	2.97*	12.4	62.0	2.54*
Shapes	11.9	59.5	2.31*	13.2	66.0	3.44***
Experiment 5B						
Tones	13.7	68.5	4.25**	13.1	65.5	3.44**
Nonwords	12.0	60.0	2.58*	12.2	61.0	3.44*

Note. Mean values reported out of 20 possible correct. T-tests are conducted with respect to chance levels. The Colors-Baseline and Tones-Baseline are the same data reported in Experiment 4 (called Visual-Baseline and Auditory-Baseline).

* $p < .05$. ** $p < .01$. *** $p < .001$.

The results for Experiments 5A and 5B were similar to those of Experiment 4. Participants were adept at learning two different sets of statistical regularities simultaneously within the same sense modality, for

shape and color sequences (Experiment 5A) and tone and nonword sequences (Experiment 5B). Performance levels in these dual-grammar conditions were no worse than learning levels with one grammar only. These results thus suggest that participants can acquire statistical regularities from two streams of information within the same sense modality, as long as the two streams differ along a major perceptual dimension. Furthermore, these results suggest that participants' knowledge of the underlying statistical structure was stimulus-specific rather than abstract.

We next look at dual-grammar learning within the same sense modality when the vocabularies lie along the same perceptual dimension. That is, we use two different sets of shapes and two different sets of nonwords.

Experiment 6: Intramodal Learning Within the Same Perceptual Dimension

The purpose of Experiment 6 was to test whether learners can learn two sets of statistical regularities when they are presented along the same perceptual “dimension” within the same sense modality. Experiment 6A incorporated two different sets of visual shapes and Experiment 6B incorporated two different sets of auditory nonwords.

Method

Subjects. For Experiment 6, 60 additional participants (10 in each condition) were recruited.

Materials. Experiment 6 incorporated the same two grammars, training and test sequences that were used in Experiments 4 and 5. Like the previous experiments, the experimental conditions employed learning under dual-

grammar conditions. Experiment 6A employed two visual vocabularies: Shape Sets 1 and 2 (Figure 4.3). Shape Set 1 was the same set of shapes used in Experiment 5A; Shape Set 2 was a new set of shapes similar in overall appearance but perceptually distinct from set 1. Experiment 6B employed the nonword vocabulary used in Experiment 5B as well as a new nonword set consisting of “tood,” “jeen,” “gens,” “tam,” and “leb.”

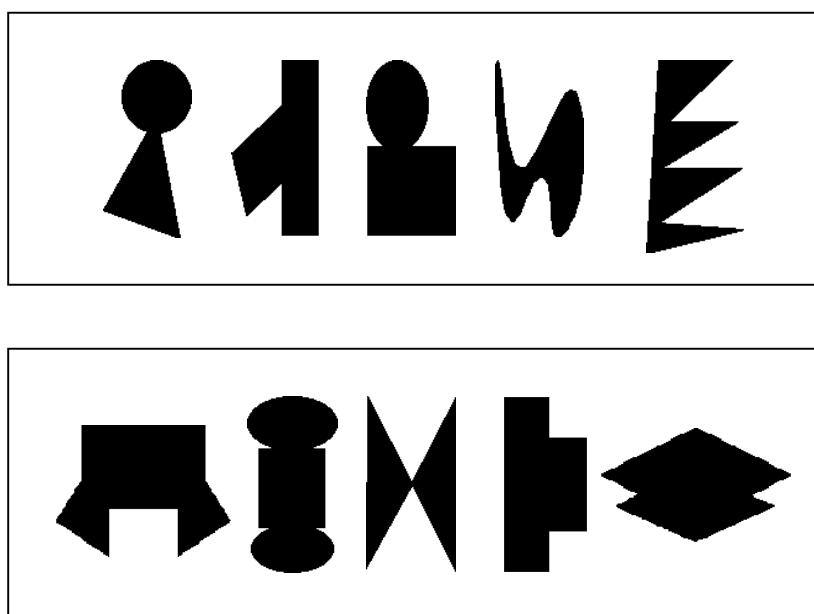


Figure 4.3. Visual vocabularies used in Experiment 6A (Shape Set 1: top; Shape Set 2: bottom).

Procedure. Participants were randomly assigned to one of six conditions, two for Experiment 6A, two for Experiment 6B, and two new single-grammar baseline conditions. The general procedure was identical to Experiment 5 except that different vocabularies were used. In Experiment 6A, one grammar was instantiated as 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 6B 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 5). In all other respects, the procedure for Experiment 6 was the same as in Experiment 5.

Results and Discussion

Mean scores and t-tests compared to chance levels are shown in Table 4.4. When exposed to two different statistically-governed streams of visual input, each with a distinct vocabulary of shapes, learners on average were 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. Thus, under dual-grammar conditions, learners showed above-chance classification performance for only one of the vocabularies/ grammars. As we remarked earlier, chance level performance could be due to either an inability to learn the underlying regularities or the result of having acquired these regularities in terms of abstract representations that do not distinguish items based on perceptual characteristics. Thus, the current data implies one of the following: 1) dual-grammar statistical learning did not occur due to perceptual

confusion of the stimuli; or 2) once learned, the knowledge of the two grammars was co-mingled because the input elements were perceptually similar. Either way, traditional, abstractive theories of AGL may have difficulty accounting for such low-level, perceptual effects.

Table 4.4

Experiment 6 Mean Performance and Tests of Significance versus Chance

Dimension	<u>Experimental (Dual-grammar)</u>			<u>Baseline (Single-grammar)</u>		
	<i>M</i>	%	<i>t</i> (9)	<i>M</i>	%	<i>t</i> (9)
Experiment 6A						
Shapes1	12.0	60.0	2.58*	13.2	66.0	3.44***
Shapes2	11.2	56.0	1.65	11.6	58.0	2.95*
Experiment 6B						
Nonwords1	10.9	54.5	1.49	12.2	61.0	3.44*
Nonwords2	12.4	62.0	6.47***	13.3	66.5	3.79**

Note. Mean values reported out of 20 possible correct. T-tests are conducted with respect to chance levels. The Shapes1-Baseline and Nonwords1-Baseline are the same data reported in Experiment 5 as Shapes-Baseline and Nonwords-Baseline.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Overall 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 contrasted 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 two-way interactions (p 's $> .4$). There was a significant three-way interaction, however, $F(3, 144) = 2.59, p = .055$, driven by better auditory-crossmodal learning for Grammar A but better visual-crossmodal learning for Grammar B. 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).

Figure 4.4 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, this outcome confirms that there was a learning decrement in Experiment 6 for intramodal learning when the two grammars were composed of vocabularies along the same perceptual dimension.

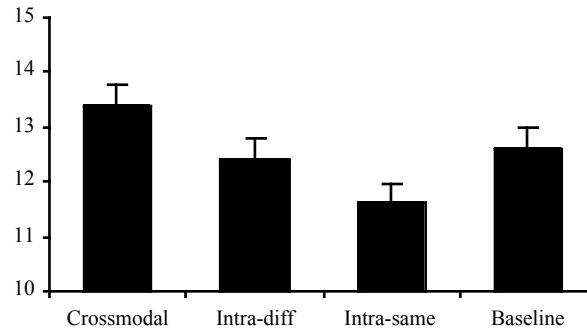


Figure 4.4. Mean test performance (out of 20) for all three experiments: Crossmodal (Experiment 4), Intramodal, different-dimension (Experiment 5), Intramodal, same-dimension (Experiment 6), and Baseline, single-grammar conditions (Experiments 4, 5, 6).

General Discussion

These experiments sought to determine the nature of acquired knowledge underlying implicit statistical learning. We distinguished between two possibilities. On the one hand, consistent with traditional information-processing approaches, it is possible that people encode the underlying structure of complex patterns in an amodal or abstract fashion. On the other hand, embodiment theories propose that the mind represents concepts by relying on modality-specific sensorimotor systems. Our data support the latter view.

Experiment 4 showed that participants 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. This result suggests that

not only was learning modality-specific, but that the underlying learning systems operate in parallel and independently of one another. Experiments 5 and 6 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 were along the same perceptual dimension, limiting statistical learning to just one of the two input streams.

This modality-dependent view of statistical learning stands in contrast to claims that learning in an AGL task may consist of modality-independent representations (Altmann et al., 1995) or abstract “rules” (Marcus et al., 1999; Reber, 1993). Some AGL studies purportedly show transfer effects across modalities, suggesting that the underlying knowledge is abstract. However, there has been considerable controversy surrounding the transfer data (e.g., Christiansen & Curtin, 1999; Marcus, 1999; Mathews, 1990; McClelland & Plaut, 1999; Redington & Chater, 1996). For example, transfer may be achieved by participants noticing the presence of low frequency illegal starting elements in the transfer set, rather than by relying on abstract knowledge per se (Tunney & Altmann, 1999). Or, participants may merely be recognizing certain patterns of repeating elements (e.g., “BDCCCCB”) and then recognizing the same repetition patterns in items with a new vocabulary (e.g., “MTVVVVM”; Brooks & Vokey, 1991; Redington & Chater, 1996). Thus it is far from clear that transfer effects reflect the operation of abstract knowledge acquired during the learning task.

Though the current data point toward modality-specificity, it is possible that human cognition relies on stimulus-specific representations for some tasks

but abstract learning for others. For example, explicit problem-solving tasks sometimes tap participants' use of abstract principles (Goldstone & Sakamoto, 2003; Reeves & Weisberg, 1994). The ability to learn abstract principles and transfer them to new domains certainly appears to be a hallmark of explicit cognition; it is much less clear, especially in light of the current data, whether this is also true for implicit learning.

In addition to the evidence for modality-specificity, the data revealed, quite remarkably, that participants can learn multiple, independent statistical regularities simultaneously. This ability makes sense when one considers that humans often attend to and process multiple, concurrent perceptual inputs at the same time, especially across different sensory modalities.¹⁴ 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.

Finally, these data speaks to the nature of the underlying cognitive mechanisms of statistical learning. It is commonly believed that statistical learning involves a single, unitary mechanism that operates over all types of input (e.g., Kirkham, Slemmer, & Johnson, 2002). However, our data may indicate that this view is inaccurate, or at least, incomplete. It is not clear how a single, amodal mechanism could afford simultaneous learning of multiple statistical regularities and keep the stimulus-specific representations independent of one another (Experiments 4 and 5). Previous research has suggested that although commonalities exist with statistical learning across

¹⁴ For instance, people can better attend to rapidly-presented sequential stimuli when one stream is auditory and the other is visual, compared to when both streams are in the same modality (Duncan, Martens, & Ward, 1997).

vision, audition, and touch, there also are important modality differences, highlighting the possible operation of separate modality-constrained subsystems (Conway & Christiansen, 2005). Such a view of statistical learning resonates both with theories of implicit sequence learning (Goschke, 1998; Keele, Ivry, Mayr, Hazeltine, & Heuer, 2003) and implicit memory (Schacter, Chiu, & Ochsner, 1993).

Implicit memory research in particular may offer insights into the nature of statistical learning mechanisms. It appears likely that both implicit statistical learning and perceptual priming are supported by something akin to perceptual fluency (Chang & Knowlton, 2004; Kinder, Shanks, Cock, & Tunney, 2003). Networks of neurons in modality-specific brain regions show decreased activity when processing items that are the same or similar in overall structure – possibly because of increased processing efficiency for that class of stimuli (Reber, Stark, & Squire, 1998; Schacter & Badgaiyan, 2001). A perceptual priming/fluency explanation for statistical learning is consistent with the stimulus-specific learning we observed in the current experiments and possibly offers an attractive possibility of unifying implicit learning and implicit memory phenomena.

Much of perception and cognition involves the use of multiple sense modalities to extract structure from temporal or spatiotemporal patterns. The current experiments suggest that the knowledge underlying such implicit statistical learning is closely tied to the sensory and perceptual features of the material itself, perhaps indicating the involvement of multiple learning subsystems, and challenging traditional abstractive and amodal theories of cognition.

References

- 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.
- Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-660.
- Barsalou, L.W., Simmons, W.K., Barbey, A.K., & Wilson, C.D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, 7, 84-91.
- Brooks, L.R. & Vokey, J.R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General*, 120, 316-323.
- 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. (1993). *Mechanisms of implicit learning: Connectionist models of sequence processing*. Cambridge, MA; MIT Press.
- 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.

- Fodor, J.A. (1975). *The language of thought*. New York: Thomas Y. Crowell.
- Glenberg, A.M. (1997). *What memory is for*. *Behavioral and Brain Sciences*, 20, 1-55.
- Goldstone, R.L. & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology*, 46, 414-466.
- Gomez, R.L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13, 431-436.
- Gomez, R.L. & Gerken, L.A. (1999). Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. *Cognition*, 70, 109-135.
- 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.
- Kinder, A., Shanks, D.R., Cock, J., & Tunney, R.J. (2003). Recollection, fluency, and the explicit/implicit distinction in artificial grammar learning. *Journal of Experimental Psychology: General*, 132, 551-565.
- Lakoff, G. (1988). Smolensky, semantics, and the sensorimotor system. *Behavioral and Brain Sciences*, 11, 39-40.
- Manza, L. & Reber, A.S. (1997). Representing artificial grammars: Transfer across stimulus forms and modalities. In D.C. Berry (Ed.), *How implicit is implicit learning?* (pp. 73-106). Oxford University Press.
- Marcus, G.F. (1999). Connectionism: With or without rules? Response to J.L. McClelland and D.C. Plaut. *Trends in Cognitive Sciences*, 3, 168-170.

- Marcus, G.F., Vijayan, S., Rao, S.B., & Vishton, P.M. (1999). Rule learning by seven-month-old infants. *Science*, 283, 77-79.
- Mathews, R.C. (1990). Abstractness of implicit grammar knowledge: Comments on Perruchet and Pacteau's analysis of synthetic grammar learning. *Journal of Experimental Psychology: General*, 119, 412-416.
- Mathews, R.C., Buss, R.R., Stanley, W.B., Blanchard-Fields, F., Cho, J.-R., & Druhan, B. (1989). The role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 15, 1083-1100.
- 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.
- Pena, M., Bonatti, L.L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298, 604-607.
- Perruchet, P. & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, 119, 264-275.
- Perruchet, P., Tyler, M.D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General*, 133, 573-583.
- Pylyshyn, Z.W. (1984). *Computation and cognition*. MIT Press.
- Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Behavior*, 6, 855-863.
- Reber, A.S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, 81, 115-119.

- Reber, A.S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118, 219-235.
- 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.
- Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General*, 125, 123-138.
- Reeves, L.M. & Weisberg, R.W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin*, 115, 381-400.
- 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.
- Schacter, D.L. & Badgaiyan, R.D. (2001). Neuroimaging of priming: New perspectives on implicit and explicit memory. *Current Directions in Psychological Science*, 10, 1-4.
- 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.
- Servan-Schreiber, E. & Anderson, J.R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 16, 592-608.
- Shanks, D.R., Johnstone, T., & Staggs, L. (1997). Abstraction processes in artificial grammar learning. *The Quarterly Journal of Experimental Psychology*, 50A, 216-252.

- Tunney, R.J. & Altmann, G.T.M. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27, 614-639.
- Tunney, R.J., & Altmann, G.T.M. (1999). The transfer effect in artificial grammar learning: Reappraising the evidence of the transfer of sequential dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1322-1333.

CHAPTER FIVE

Summary and Discussion

The studies presented in this thesis provide groundwork for a perceptual theory of implicit statistical learning. In contrast to traditional perspectives on artificial grammar learning (AGL), these three studies emphasize the importance of sense modality and stimulus characteristics in learning. The data can be summarized with three main points regarding implicit learning: 1) learning is affected by auditory-temporal and visual-spatial constraints; 2) learning is affected by auditory-final, visual-initial, and tactile-initial biases; 3) and learning is mediated by independent, stimulus-specific representations. After considering each in turn, we present a preliminary outline of a perceptual theory of implicit learning, supported in part by drawing upon principles of perceptual learning. Finally, we propose a rethinking of the standard “systems of memory” view, replacing it with a modality-constrained view of cognition.

Auditory-Temporal and Visual-Spatial Constraints

One form of modality constraint observed in the experiments was the presence of auditory-temporal and visual-spatial emphases in the processing of complex patterns (Chapters 2, 3, 4). Consistent with previous research in memory and perception, the data highlight the fact that audition and vision are differentially adept at picking up information distributed in the dimensions of time and space, respectively (Collier & Logan, 2000; Friedes, 1974; Kubovy, 1988; Mahar, Mackenzie, & McNicol, 1994; O’Connor &

Hermelin, 1978; Penney, 1989). Although research had demonstrated that such constraints affect the learning of fixed sequential patterns, (Handel & Buffardi, 1969), these constraints had not been investigated in the realm of statistical learning. In Chapter 2 (Conway & Christiansen, 2005), we took measures to ensure that the temporal/spatial constraints were not due merely to differences in the ability for each sense modality to perceive the actual stimulus elements. Thus, the observed effects appear to be influencing the acquisition of the regularities existing among the relationships of elements, rather than in detecting the individual elements themselves.

Because these temporal/spatial biases affect not just implicit statistical learning but also a wide variety of other cognitive processes, this suggests the presence of global constraints affecting processing within each sense modality. In Chapter 3 we suggested that one possible cause could arise from the differences in how stimulus energies are encoded across the senses. That is, whereas the auditory system is organized tonotopically, the visual system is organized retinotopically. Sound is a pressure wave that is represented at the tympanic membrane as a one-dimensional temporal waveform. Therefore, in order to detect changes in the sound wave, the auditory system necessarily must be sensitive to these temporal dynamics. A visual image is a two-dimensional spatially distributed pattern encoded by the retina and therefore space is much more inherent in the processing of a visual image. Likewise, whereas the auditory system breaks down incoming sound waves into their constituent frequency components, leading to a tonotopic organization in auditory cortex, the visual system maintains a spatial/retinotopic organization throughout. Thus, from very early on in the processing stream, differences in

the way the stimulus energies are encoded likely affect all subsequent processing done by each sense modality, including statistical learning.

An alternative, or perhaps complementary, explanation for the auditory-temporal and visual-spatial biases is suggested by embodiment theory. A.M. Glenberg (personal communication, February 23, 2005) proposed that visual statistical learning might be mediated through eye movements. The suggestion is that the underlying regularities are encoded by the ocular motor system – as long as the regularities occur spatially. If the regularities exist not in the dimension of space but instead in purely temporal patterns, as in Chapter 3's visual-temporal condition, then eye movements cannot participate in learning and performance declines. This is an interesting hypothesis that needs further empirical investigation.

Another possible explanation for the auditory-temporal and visual-spatial biases, to be discussed later in the chapter, comes from perceptual learning mechanisms. Next, we consider the second modality constraint that was observed in the experiments.

Auditory-Final, Visual-Initial, and Tactile-Initial Biases

A second modality constraint was seen in that each sense modality was especially attuned to extracting regularities from the initial or final parts of stimuli (Chapters 2 & 3). That is, participants' endorsements of novel items in the test phase were more influenced by the fragment information contained in final positions when the input was auditory, whereas endorsements for visual or tactile material were more dependent upon fragment-initial information. Nowhere else, to our knowledge, has such an effect been demonstrated in AGL, statistical learning, or implicit learning more generally. Intriguingly,

there is a similar effect seen in the shape of serial position curves in explicit recall. It has long been known that a recency effect exists for auditory over visual material; that is, people show better recall for auditory input compared to visual input for the final one or two elements in a list (Crowder, 1986; Engle & Mobley, 1976). More recently, evidence has revealed a visual primacy effect, with recall better for visual material compared to auditory items for the first few elements in a list (Beaman, 2002; Wright, 2002). Wright (2002) observed the visual primacy effect in monkeys when recall was tested at relatively long delays (10 seconds or more). Similarly, Beaman (2002) obtained the effect when human participants were instructed to recall the second half of a list before the first half. Thus, in both cases, it appears that the visual serial recall advantage for initial items is seen only under conditions in which the recall occurs after a substantial delay.

Likewise, the test phase in an AGL task occurs many minutes after exposure to the first training stimulus. Thus, sensitivity to elements (or fragments) in the initial positions of a visual sequence may only be measured under appropriate test conditions. If so, then this suggests that common processes may underlie, or constrain, learning and memory in both implicit and explicit tasks. Finding common properties for the encoding of material under implicit and explicit conditions may force a re-thinking of theories of implicit (i.e., nondeclarative) and explicit (i.e., declarative) cognition. The most accepted view is that implicit and explicit learning and memory are mediated by entirely separate cognitive and neural systems, with different properties characterizing each (Tulving & Schacter, 1990). It is too early to tell to what extent the current data alters this standard view. There exist at least three possibilities: the auditory-final, visual-initial, and tactile-initial biases in both

implicit and explicit cognition may signify that both may be part of the same memory system, relying on a single set of underlying mechanisms; alternatively, there may be two separate systems that have common properties. A third possibility is that “implicit” learning tasks tap explicit processes. That is, participants may be explicitly recalling fragments of elements from the training phase when classifying novel strings, in which case recall of such fragments is constrained by serial position effects.

Independent, Stimulus-Specific Representations

In Chapter 4, we presented data showing that learning in AGL is stimulus-specific rather than abstract or amodal. Furthermore, the acquired knowledge existed independently for each sense modality or perceptual dimension. That is, learners acquired the statistical regularities from two input streams in parallel – either visually and auditorily or along two different dimensions within the same modality – and the representations were not confused with one another.

The issue of stimulus-specific versus abstract representations speaks to the heart of theories of AGL as well as to the wider field of cognition. In terms of cognition as a whole, our data is more consistent with an embodiment approach (Barsalou, 1999; Glenberg, 1997) rather than one that compartmentalizes perceptual processing from amodal cognition (Fodor, 1975; Pylyshyn, 1999). In terms of implicit learning specifically, it has been proposed that the acquired knowledge consists of abstract, algebraic-like rules (Reber, 1993; Marcus, Vijayan, Rao, & Vishton, 1999) because of data apparently showing that participants can transfer their acquired knowledge across different perceptual domains or modalities (e.g., Brooks & Vokey, 1991; Reber,

1969; Tunney & Altmann, 2001). However, much of the transfer data can be explained due to 1) not employing proper methodological controls; or 2) by participants adopting explicit strategies during the test task to induce mappings between vocabularies (Redington & Chater, 1996). Both possibilities remove the need for postulating acquired abstract knowledge.

The data from Chapter 4 support a non-abstractive account of AGL, with participants learning statistical structure specific to the sense modality or perceptual dimension in which it occurred. If participants had been learning the underlying abstract structure of the two input streams in our cross-over design, then they should have exhibited chance level performance. Our results show some conceptual consistency with an “exemplar-based” view of implicit learning (Brooks & Vokey, 1991; Pothos & Bailey, 2000; Vokey & Brooks, 1992), which proposes that participants make classification judgments based on overall similarity to observed training items, rather than knowledge of underlying abstract rules.¹⁵

Not only were the acquired representations stimulus-specific, but they appeared to exist separately and independently of one another. That is, participants were exposed to two statistically-governed streams, and tested on their ability to classify novel sequences. These novel sequences contained exemplars from each of the two grammars, but all test items were instantiated with one of the vocabularies only, counter-balanced across participants. The fact that participants in Experiments 1 and 2 showed high performance levels – statistically no different than single-grammar baseline conditions – means

¹⁵ Because it has been established that AGL also consists of encoding the presence of chunks of elements (i.e., fragments; Perruchet & Gallego, 1997; Perruchet & Pacteau, 1990; 1991), it is likely that learning involves both exemplar and fragment-based processes, or a combination of both.

that they learned the two regularities in parallel, keeping each stimulus domain separate from the other. For example, on average, learners classified a visual stimulus as “legal” only if it matched the statistical regularities present in the visual training items.

Several other cognitive theories are relevant to this issue of independent, stimulus-specific representations. First, as mentioned in Chapter 3, Penney (1989) outlined a “separate streams” hypothesis in which verbal short-term recall is mediated by modality-dependent processing streams, each taking in different types of input as well as having different processing capabilities. There are obvious similarities between Penney’s “separate streams” and Baddeley and Hitch’s (1974; 1994; Baddeley, 1992) multicomponent model of working memory, which also contains modality-dependent processing loops. How the Penney / Baddeley account of short-term memory relates to implicit statistical learning is discussed further below. It will be suggested that both short-term memory and implicit learning operate within a wider, modality-specific cognitive system.

Second, Keele, Ivry, Mayr, Hazeltine, and Heuer (2003) proposed a neurocognitive model of sequence representation, based mainly on data from the serial reaction-time task (Nissen & Bullemer, 1987). Their model includes two separate pathways, dorsal and ventral, differing in neural structures, in the type of input that is processed, and in access to awareness. In brief, the dorsal stream is implicit and encodes unidimensional regularities whereas the ventral stream is both implicit and explicit and can associate elements across modalities or dimensions. The Keele et al. (2003) model, although based on data from a different experimental paradigm, may be applicable to the perceptual-based sequence learning explored in the current experiments.

Specifically, Keele's unidimensional "modules" may serve stimulus-specific learning that was observed in this thesis. Furthermore, their description of the multidimensional processing system lays the groundwork for exploring implicit statistical learning that involves input occurring across different perceptual modalities or dimensions. Keele et al. propose that learning cross-modal or cross-dimensional patterns requires active attention and awareness. Thus, this prediction can be explored within the context of perceptual sequence learning using the AGL task.

Third, a popular view of implicit memory holds that it is composed of separate, modality-specific subsystems, such as an auditory and a visual perceptual representation system (Schacter, 1994). These subsystems mediate priming phenomenon, in which people have a better ability to identify or produce items following prior exposure to those items without conscious recollection of the exposure (Tulving & Schacter, 1990). The modality-specific subsystems are inferred based on evidence that priming is significantly reduced when the modality or surface features are changed between study and test (Schacter, 1987). Additionally, there appears to be very little interaction between the visual and auditory priming subsystems (David & Hirshman, 1998). The existence of separate priming subsystems is also supported by neuroimaging data showing decreased activity in modality-specific brain areas during a priming task; for instance, within-modal visual priming is accompanied by decreased activity in occipitotemporal extrastriate cortical regions (Schacter, Dobbins, & Schnyer, 2004). As we suggested in Chapter 4, something akin to perceptual priming may also support classification tasks in implicit statistical learning. Some support for this view comes both from behavioral (Chang & Knowlton, 2004; Kinder, Shanks, Cock,

& Tunney, 2003) and neuroimaging studies (Reber, Stark, & Squire, 1998), which emphasize the role of perceptual processing in implicit learning.

Implicit Learning as Perceptual Learning

To sum up so far, our data reveal three main points. Implicit statistical learning shows: 1) auditory-temporal and visual-spatial biases; 2) auditory-final, visual-initial, and tactile-initial biases; 3) evidence for independent, stimulus-specific processing. We suggest that these modality-specific constraints on statistical learning point toward the existence of separate underlying learning mechanisms, each with different properties and characteristics, and each having the ability to operate independently and in parallel to one another.

Based on all of this evidence, we suggest that implicit statistical learning arises from and is continuous with perceptual processing. Such a view is consistent with embodiment theories (Barsalou, 1999; Glenberg, 1997; see also Goldstone & Barsalou, 1998) that argue that cognition as a whole is supported by and made up of perceptual (and motor) mechanisms. This raises the question, what exactly are the perceptual mechanisms that support implicit statistical learning? One possible mechanism already mentioned is perceptual fluency (a.k.a. priming), in which perceptual processing becomes increasingly efficient after exposure to particular items or item types.

Other insights regarding perceptual mechanisms may come from perceptual learning phenomena. Perceptual learning involves changes to an organism's perceptual system so that it can respond better to its environment (Goldstone, 1994). Superficially, there appear to be similarities between perceptual learning and implicit learning. For instance, participants' verbal

reports in our experiments suggest that their perception of the stimuli has changed after exposure to the training exemplars: participants often claim to have made the test responses based on what “looked” or “sounded” right. One hint of the relevance of perceptual learning to statistical learning comes from Hall (1991), who suggested links between associative learning – which in turn possibly mediates implicit learning (Cleeremans, Destrebecqz, & Boyer, 1998) -- and perceptual learning. In fact, the Gibsons’ perceptual learning approach (Gibson & Gibson, 1955) formed the basis for Reber’s (1967) first study of implicit learning. Reber hypothesized that learning the regularities of language-like structure may occur through something akin to the Gibsons’ idea of “differentiation,” where the learner becomes more sensitive to the information inherent in the environment due to mere exposure to the relevant stimuli. Aside from this early remark, there have been virtually no explicit connections

made between implicit and perceptual learning.¹⁶ One exception is a recent talk given at Cornell University (Goldstone, 2002) in which it was proposed that statistical learning and perceptual learning both rely on the same mechanisms. Below, I expand upon this idea, using the four perceptual learning mechanisms outlined by Goldstone (1994): attentional weighting, stimulus imprinting, differentiation, and unitization.

Attentional weighting

Attentional weighting modifies an organism’s attention, increasing it to important perceptual dimensions or features or decreasing it to irrelevant

¹⁶ It may be that Reber’s (1969) subsequent paper showing transfer of implicit learning between different letter sets discouraged the notion that perceptual learning mechanisms are synonymous with implicit learning. That is, if one believes that the transfer data point to abstract knowledge, then it is difficult to see a role for perceptual learning.

dimensions or features. Certain dimensions or stimulus features acquire “distinctiveness” because they become important to the organism in some way. Goldstone (1998) suggested that one example of attentional weighting is categorical perception (Liberman et al., 1957), where between-category features are emphasized over within-category features.

Attentional weighting may be the mechanism by which our learners showed the auditory-final, visual-initial, and tactile-initial biases. An attentional weighting account would suggest that organisms develop (phylogenetically or ontogenetically) greater sensitivity for the beginning or ending of input sequences in the different senses. It is not clear whether the weighting is innate or has a developmental progression. Either way, presumably these attentional biases would have developed because they confer a functional advantage. Perhaps having different shaped serial position curves is advantageous because where learning and memory for one sense modality is strongest (auditory-final), the other senses are weakest (vision, touch), and vice-versa (Wright, 2002).

An attentional weighting account might also explain the auditory-temporal and visual-spatial differences. Perhaps organisms’ attention to the dimensions of space and time change through exposure of the auditory and visual world. That is, attention toward auditory-temporal and visual-spatial information is increased, whereas attention to auditory-spatial and visual-temporal information is decreased. This account differs from the account we offered earlier which suggested these constraints might arise from the physical make-up of the sensory systems. One way to distinguish between the two accounts would be to explore whether young infants have these modality constraints. If they do, it might imply that it occurs as a product of the

physical architectures of the sensory systems as opposed to developing from experience to sensory input.

Stimulus imprinting

Related to attentional weighting, stimulus imprinting refers to the development of specialized functional “detectors” to detect important or frequently repeating stimuli.¹⁷ Interestingly, Goldstone remarks that stimulus imprinting may be the mechanism that explains implicit memory phenomenon. In terms of implicit statistical learning, detectors could be developed for whole stimuli (i.e., entire training sequences) or parts of stimuli (i.e., frequently co-occurring elements), neatly coinciding with exemplar (Vokey & Brooks, 1992) and fragment-based (Perruchet & Pacteau, 1990) theories of AGL. According to Goldstone, the creation of fragment/part detectors allows the organism to develop new “building blocks” for representing stimuli, thus allowing for the flexible description of novel objects and stimuli. In the case of statistical learning, these building blocks correspond to frequently co-occurring elements (fragments or chunks), allowing learners to generalize their knowledge to novel stimuli in an AGL test task.

Unitization and Differentiation

Unitization is the mechanism by which a single perceptual chunk is created out of separate stimulus elements that reliably co-occur.

Differentiation refers to the mechanism in which perceptual elements that had originally been fused together become separated. As Goldstone (2002)

¹⁷ Goldstone (1994) uses the term “detector” to refer to any abstract/cognitive device or process.

suggested, unitization and differentiation are likely candidates for implicit statistical learning. Elements that reliably co-occur in AGL tasks can be represented as chunks or fragments (Perruchet & Pacteau, 1990). Similarly, differentiation might help separate elements that weakly co-occur. These two mechanisms thus are opposites but operate in concert. Elements that do not occur together reliably become increasingly differentiated at the same time that highly co-occurring elements become unitized.

One or all of these four mechanisms may underlie aspects of implicit statistical learning. In comparing these mechanisms to posited theories of AGL, a perceptual learning account differs from abstractive, rule-learning theories but appears to share similarities with associative/fragment based accounts (Cleeremans, Destrebecqz, & Boyer, 1998; Perruchet & Pacteau, 1990) as well as exemplar theories (Vokey & Brooks, 1992). Associative/fragment accounts posit that frequently co-occurring elements are associated or “chunked” together. This account shares obvious similarities with unitization, for example. Exemplar models propose that whole items are encoded during the training task and that test items are then compared with these stored exemplars. Thus, something akin to stimulus imprinting appears similar to the exemplar approach.

Despite the similarities, a perceptual learning account differs from associative and exemplar theories because it posits that perception itself changes. These changes presumably operate at a relatively “low” level in the information stream, suggesting that the locus of learning may be in primary sensory cortices (Fahle, 2002; Goldstone, 1998).

Taking all this into consideration, the groundwork is now in place for constructing a perceptual theory of implicit statistical learning.

Toward a Perceptual Theory of Implicit Learning (and Beyond)

Here we outline the beginnings of a perceptual theory of implicit statistical learning. We begin by proposing four characteristics of statistical learning. We then place statistical learning within the wider context of perception and cognition and go on to present a new way of thinking about systems of perception, learning, and memory.

Given the evidence presented in this thesis, we propose that implicit statistical learning has the following four characteristics.

- 1) Learning is mediated by multiple, modality-specific (visual, auditory, tactile, etc.) mechanisms.
- 2) Each learning mechanism operates independently and potentially in parallel with the others, although some interaction can occur.
- 3) Each learning mechanism has unique properties that constrain learning (e.g., the auditory system specializes in handling temporal input and is also biased to encode material in list-final positions).
- 4) Each mechanism operates using similar computational principles.

These principles may be a combination of those involved in implicit memory (priming) and perceptual learning.¹⁸

A question that naturally follows is how do these modality-specific implicit learning mechanisms relate to the rest of cognition. One prominent view holds that learning, memory, and cognition can be subdivided into multiple components, in a more or less hierarchical manner (Schacter &

¹⁸ It will be necessary for future research to begin to distinguish between these alternate mechanisms and the role they play in implicit statistical learning.

Tulving, 1994; Squire & Zola, 1996). These components include working memory, declarative memory, nondeclarative memory, priming, etc. (see Figure 5.1). Some of these “systems” in turn contain modality-specific subsystems. We propose an alternative way of depicting the components of cognition that turns the standard view on its head. Instead of positing separate memorial systems that are in turn composed of modality-specific subsystems, we propose that cognition consists of separate sensorimotor cognitive systems, each composed of a suite of inter-related perceptual, memorial, and cognitive functions. This view recognizes the major, if not primary, role that perception plays, and the continuous nature of perception with cognition and action (Barsalou, 1999; Fuster, 1997; Lloyd, 2000; Mesulam, 1998).

The proposed cognitive architecture is organized by the primary sense modalities. Figure 5.2 shows the Visual, Auditory, and Tactile Cognitive Systems. The purpose of each of these systems is to detect, encode, and represent the immediate and past environment in order to prepare the organism for future action. Thus, each of these systems serve a number of related processes, such as implicit memory, implicit learning, long-term memory, and language. We hypothesize that these within-modality processes rely upon and overlap with one another, rather than being discrete mechanisms as is generally assumed.

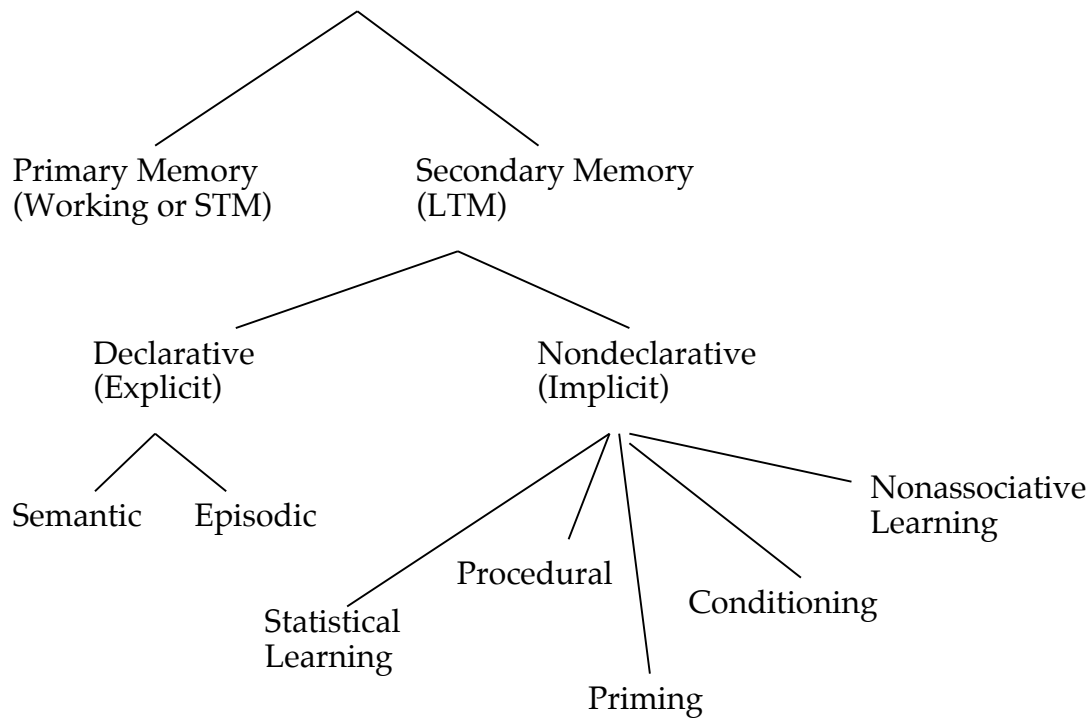


Figure 5.1. A prominent view of the systems of memory and cognition. Adapted and modified from Squire and Zola (1996).

Note that in Figure 5.2, the outlines of each system are illustrated with dashed lines, meant to imply that the boundaries are not sharp and rigid. Thus, Fodorian “information encapsulation” (Fodor, 1975) is not a characteristic of this model. There is a large body of research demonstrating crossmodal interactions in perception and cognition (e.g., Calvert, 2001; Driver & Spence, 2000; McGurk & MacDonald, 1976; Meredith, 2002). The diagram represents these interactions with the connecting arrows. The specific manner and properties of such interactions are not within the scope of this thesis. Likewise, a “multimodal” system or systems likely exist but are beyond the scope of the current discussion.

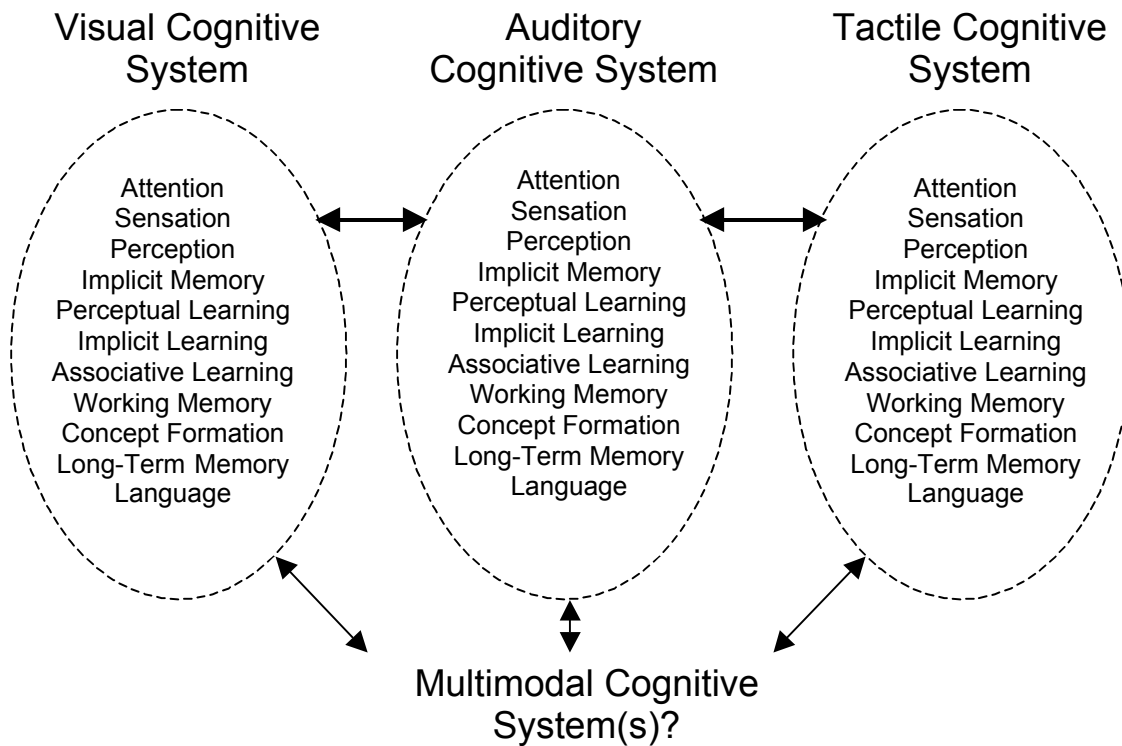


Figure 5.2. Proposed modality-constrained view of cognition. Each perceptual/cognitive system is composed of a suite of inter-related perceptual, memorial, and cognitive functions. The list of processes listed within each system is not meant to be exhaustive. Also note that because this thesis has been focused exclusively on the visual, auditory, and tactile modalities, we exclude olfactory, taste, and motor cognitive systems from the figure, while recognizing that they surely exist.

Each perceptual/cognitive system is proposed to have certain modality-specific global constraints that affect all of its processes. In this way, the model may help explain why the same modality constraints – e.g., visual-spatial and auditory-temporal biases – are seen in apparently disparate cognitive processes, including implicit learning, working memory, long-term memory, and perception.

Below, we list and briefly describe some of the proposed functions that each modality-constrained cognitive system might include.

Attention

Both anecdotal and empirical evidence suggests that attention has a modality-specific component. That is, it is easier to attend to two stimuli simultaneously when they occur in two different sense modalities rather than in the same one (Duncan, Martens, & Ward, 1997; Treisman & Davies, 1973). Furthermore, attention appears to be implemented neurally by a modality-specific competition and suppression mechanism (Duncan, 1999). For example, neurons corresponding to visual objects in visual cortex show suppressed responses when the objects are being ignored, whereas attended objects are strongly represented. This evidence suggests that attention may arise out of the workings of separate modality-specific cognitive systems.

Perceptual Representation

Borrowing from Schacter's terminology (1994), we use the term "perceptual representation" to refer to not just perception and implicit memory (priming), but perceptual learning and implicit learning, as well. As already discussed and outlined, we suggest that these four phenomena may be variations of the same overall perceptual processing mechanisms.

Associative Learning

Associative learning has an intimate connection with implicit statistical learning in that it appears to be a mechanism for uncovering environmental regularities (Cleeremans et al., 1998; Ryder & Fodorov, 2001). However, although associative learning has long been believed to be an important aspect of learning and memory, it has mostly been considered unable to account for perceptual learning phenomenon (Goldstone, 1998; though also see Hall, 1991 for an effort to unite associative and perceptual learning). Regardless, associative learning is a core process in both cognitive and neural function

(Hebb, 1949; Fuster, 1997). It seems likely that this learning process supports perceptual and cognitive functions within and across each modality-specific system. For instance, learning an association between two visual stimuli likely involves links established between visual perceptual representations, whereas learning an association between a visual and auditory stimulus necessitates cross-modal links between the two systems.

Working Memory and Serial Recall

Working memory is commonly believed to consist of separate, modality-specific processing streams, such as the phonological loop and visuospatial sketchpad (Baddeley & Hitch, 1994; Penney, 1989). These modality-specific streams appear to be mediated by portions of the prefrontal cortex that have corresponding links with the areas of the brain handling the perception of phonological and visuospatial information (Goldman-Rakic, 1999). Thus, working memory appears to naturally arise out of perceptual processes.

Postulating working memory as arising out of modality-specific perceptual processing dovetails nicely with MacDonald and Christiansen's (2002; Christiansen & MacDonald, 1999) proposal that working memory emerges from interactions between experience and domain-specific processing architectures. Thus, on the present account, there is not a separate working memory resource; instead, it is the consequence of the functioning of each modality-specific system.

Long-Term Memory, Concept Formation, and Language

Barsalou (1999) proposed that long-term memory and conceptual knowledge are derived from the same neural and cognitive mechanisms underlying perception. Simply, a perceptual state arises in modality-specific

sensorimotor systems. These perceptual representations can then be stored in long-term memory, retaining their perceptual characteristics. These modality-specific “symbols” can then be merged and recombined dynamically to create complex multimodal concepts that are less tied to direct perceptual experience, though still grounded in it.

Likewise, Pulvermuller (2001) presented evidence that the brain represents the meaning of words via functional “webs,” networks of neurons distributed throughout the brain in regions corresponding to the perceptual and motor properties of the word in question. Thus, words with strong visual associations (i.e., animal names) correspond to greater brain activity in occipital cortex whereas words with strong action connotations (i.e., tool names) result in greater activity in premotor cortex. This provides strong evidence that meaning in language is represented in a modality-specific manner.

Conclusion

This model offers a framework for unifying perception and cognition while simultaneously reducing the number of various memorial systems that have been proposed. It also suggests how different independent cognitive and perceptual processes may be more functionally and mechanistically similar than generally realized. In fact, areas that are traditionally entirely separate research endeavors – such as implicit learning, implicit memory, and perceptual learning – may be aspects of the same overall function.

This framework also offers novel predictions regarding cognition. For instance, the processes within each modality-specific system ought to be very closely related and dependent upon one another. Thus, it is predicted that working memory and implicit learning, for example, which have seldom been

investigated together, interact closely. Predictions such as this one are not borne out if one takes the standard memory systems view (Figure 5.1).

In summary, this dissertation has presented evidence for modality constraints affecting implicit statistical learning. We have suggested that these modality differences arise from the operation of separate, modality-specific cognitive systems, each with different processing constraints. These modal systems have a common purpose: to provide information regarding the various sensory domains in order to prepare the organism for future action. As such, these systems encompass the range of perceptual and cognitive processes that are studied in human and non-human animals. Unanswered questions remain, of course, such as to what extent these systems interact with one another. But it is proposed that this modality-constrained account of cognition may yield new insights into not only how organisms extract structured information from the environment, but also into the very nature of mind and behavior.

References

- Baddeley, A. (1992). Working memory. *Science*, 255, 556-559.
- Baddeley, A.D. & Hitch, G.J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation*, (Vol.8, pp.47-90). San Diego, CA: Academic Press.
- Baddeley, A.D. & Hitch, G.J. (1994). Developments in the concept of working memory. *Neuropsychology*, 8, 485-493.
- Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-660.
- Beaman, C. P. (2002). Inverting the modality effect in serial recall. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 55A, 371-389.
- Brooks, L.R. & Vokey, J.R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General*, 120, 316-323.
- Calvert, G.A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, 11, 1110-1123.
- 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. & MacDonald, M.C. (1999). Fractionating linguistic working memory: Even in pebbles, it's still a soup stone. *Behavioral and Brain Sciences*, 22, 97-98.
- Collier, G.L. & Logan, G. (2000). Modality differences in short-term memory for rhythms. *Memory & Cognition*, 28, 529-538.

- 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.
- Crowder, R. G. (1986). Auditory and temporal factors in the modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2, 268-278.
- David, P. & Hirshman, E. (1998). Dual-mode presentation and its effect on implicit and explicit memory. *The American Journal of Psychology*, 111, 77-87).
- Driver, J. & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology*, 10, R731-R735.
- Duncan, J. (1999). Attention. In R.A. Wilson & F.C. Keil (Eds.), *The MIT Encyclopedia of the Cognitive Sciences* (pp. 39-41). Cambridge, MA: MIT Press.
- Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, 387, 756-757.
- Engle, R. W., & Mobley, L. A. (1976). The modality effect: What happens in long-term memory? *Journal of Verbal Learning and Verbal Behavior*, 15, 519-527.
- Fahle, M. (2002). Perceptual learning: Gain without pain? *Nature Neuroscience*, 5, 923-924.
- Fodor, J.A. (1975). *The language of thought*. New York: Thomas Y. Crowell.
- Freides, D. (1974). Human information processing and sensory modality: Cross-modal functions, information complexity, memory, and deficit. *Psychological Bulletin*, 81, 284-310.
- Fuster, J.M. (1997). Network memory. *Trends in Neurosciences*, 20, 451-459.
- Gibson, E. & Gibson, J.J. (1954). Perceptual learning: Differentiation or enrichment? *Psychological Review*, 62, 32-41.

- Glenberg, A.M. (1997). What memory is for. *Behavioral and Brain Sciences*, 20, 1-55.
- Goldman-Rakic, P. (1999). Neural basis of working memory. In R.A. Wilson & F.C. Keil (Eds.), *The MIT Encyclopedia of the Cognitive Sciences* (pp. 890-894). Cambridge, MA: MIT Press.
- Goldstone, R.L. (2002). *Creating perceptual representations that recreate the world*. Talk given at the Cornell University Cognitive Studies Symposium: Statistical Learning Across Cognition, Ithaca, NY, April.
- Goldstone, R.L. (1998). Perceptual learning. *Annual Review of Psychology*, 49, 585-612.
- Goldstone, R.L. & Barsalou, L.W. (1998). Reuniting perception and conception. *Cognition*, 65, 231-262.
- Hall, G. (1991). *Perceptual and associative learning*. Oxford University Press.
- Handel, S., & Buffardi, L. (1969). Using several modalities to perceive one temporal pattern. *Quarterly Journal of Experimental Psychology*, 21, 256-266.
- Hebb, D.O. (1949). *The Organization of behavior: A Neuropsychological theory*. New York: John Wiley & Sons.
- 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.
- Kinder, A., Shanks, D.R., Cock, J., & Tunney, R.J. (2003). Recollection, fluency, and the explicit/implicit distinction in artificial grammar learning. *Journal of Experimental Psychology: General*, 132, 551-565.
- Kubovy, M. (1988). Should we resist the seductiveness of the space:time::vision:audition analogy? *Journal of Experimental Psychology: Human Perception and Performance*, 14, 318-320.

- Lieberman, A.M., Harris, K.S., Eimas, P.D., Lisker, L., & Bastian, J. (1957).
Journal of Experimental Psychology, 61, 379-388.
- Lloyd, D. (2000). Terra cognita: From functional neuroimaging to the map of the mind. *Brain & Mind*, 1, 93-116.
- MacDonald, M.C. & Christiansen, M.H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996).
Psychological Review, 109, 35-54.
- Mahar, D., Mackenzie, B., & McNicol, D (1994). Modality-specific differences in the processing of spatially, temporally, and spatiotemporally distributed information. *Perception*, 23, 1369-1386.
- Marcus, G.F., Vijayan, S., Rao, S.B., & Vishton, P.M. (1999). Rule learning by seven-month-old infants. *Science*, 283, 77-79.
- McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
- Meredith, M.A. (2002). On the neuronal basis for multisensory convergence: A brief overview. *Cognitive Brain Research*, 14, 31-40.
- Mesulam, M.-M. (1998). From sensation to cognition. *Brain*, 121, 1013-1052.
- Nissen, M.J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, 19, 1-32.
- O'Connor, N. & Hermelin, B. (1978). *Seeing and hearing and space and time*. New York: Academic Press.
- Penney, C.G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, 17, 398-422.
- Perruchet, P., & Gallego, J. (1997). A subjective unit formation account of implicit learning. In D.C. Berry (Ed.), *How implicit is implicit learning?* (pp. 124-161). Oxford: Oxford University Press.

- Perruchet, P., & Pacteau, C (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, 119, 264-275.
- Perruchet, P., & Pacteau, C. (1991). Implicit acquisition of abstract knowledge about artificial grammar: Some methodological and conceptual issues. *Journal of Experimental Psychology: General*, 120, 112-116.
- Pothos, E.M., & Bailey, T.M. (2000). The role of similarity in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 847-862.
- Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, 5, 517-524.
- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case of cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22, 341-423.
- Reber, A.S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, 81, 115-119.
- Reber, A.S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford: 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.
- Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General*, 125, 123-138.
- Schacter, D.L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology, Learning, Memory, & Cognition*, 13, 501-518.

- Schacter, D.L. (1994). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. In D.L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 233-268). Cambridge, MA: MIT Press.
- Schacter, D.L., Dobbins, I.G., & Schnyer, D.M. (2004). Specificity of priming: A cognitive neuroscience perspective. *Nature Reviews Neuroscience*, 5, 853-862.
- Schacter, D.L. & Tulving, E. (Eds.). (1994). *Memory systems 1994*. Cambridge, MA: MIT Press.
- Squire, L.R. & Zola, S.M. (1996). Structure and function of declarative and nondeclarative memory functions. *Proceedings of the National Academy of Sciences, USA*, 24, 13515-13522.
- Treisman, A.M. & Davies, A. (1973). Divided attention to ear and eye. In S. Kornblum (Ed.), *Attention and Performance IV* (pp. 101-117). London: Academic Press.
- Tulving, E. & Schacter, D.L. (1990). Priming and human memory systems. *Science*, 247, 301-306.
- Tunney, R.J. & Altmann, G.T.M. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27, 614-639.
- Vokey, J.R., & Brooks, L.R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 328-344.
- Wright, A. A. (2002). Monkey visual and auditory memory. In S. B Fountain, M. D. Bunsey, J. H. Danks, & M. K. McBeath (Eds.), *Animal cognition and sequential behavior: Behavioral, biological, and computational perspectives* (pp. 69-90). Boston: Kluwer Academic.