5-2013

Associative Processes in Statistical Learning: Paradoxical Predictions of the Past

Jennifer Patricia Provyn
Syracuse University

Abstract

The ability to process sequences of input and extract regularity across the distribution of input is fundamental for making predictions from the observed past to the future. Prediction is rooted in the extraction of both frequency- and conditional statistics from the distribution of inputs. For example, an animal hunting for food may consistently return to a particular area to hunt if relative to all other areas visited, that area has the highest frequency of prey. In contrast, humans asked to predict the next word in a sentence must make a prediction based upon higher-order regularities rather than simple frequency statistics (the most frequent word in the English language is *the*). The Serial Reaction Time (SRT) task, a model for studying sequential behavior, is used to quantify sensitivity to sequential constraints present in structured environments (Nissen & Bullemer, 1987). The SRT task requires Ss to make a unique response to each individually presented element from a sequence of elements. The statistics of SRT sequences, such as the relative frequency of elements and simple pairwise associations between elements, can be controlled to create dependencies that can only be predicted by learning higher-order associations. Sensitivity to the sequential constraints present in the structured input is demonstrated through differences in reaction time to elements that are, and are not, predictable based upon the statistics of the input environment. Sensitivity to statistical regularity in the environment is also a critical dimension of various episodic learning methodologies. Graded associations have been demonstrated among elements extending in both forward and backward directions in episodic memory tasks, and are suggested to reflect a gradient of the underlying structural relationships among the study elements. Graded associations are beneficial to the extent that they increase the probability of recalling sequence elements. However, unlike free and serial recall tasks, backward associations, and remote associations in general, are anti-predictive in the SRT task. The formation of associations beyond the immediately predictive element in prediction tasks could be suggestive of a ubiquitous underlying associative mechanism, which universally gives rise to graded contiguity effects,

regardless of the specific application (Howard, Jing, Rao, Provyn, & Datey, 2009). The following experiment employed a probabilistic SRT task to quantify sensitivity to immediately backward, backward-remote, and forward-remote associations. Ss were presented sequences of elements probabilistically sampled from an underlying ring-structure, with the dependent measure Ss' reaction time to elements that either followed, or deviated from, the structure. Results from the SRT task indicated that Ss demonstrated a robust backward association, as well as evidence for forward-graded associations. Moreover, in an explicit test of sequence knowledge, while Ss did not generate the probabilistic statistics from the structured learning environment, Ss did generate a statistically significant amount of backward-transitions, relative to other remote-backward transitions. The graded associations that were formed beyond the immediately predictive element in the prediction task provide evidence that a similar mechanism that mediates episodic learning may also mediate statistical learning. Backward and graded associations may be explained by a ubiquitous underlying associative mechanism, which universally gives rise to graded contiguity effects, regardless of the specific application.

ASSOCIATIVE PROCESSES IN STATISTICAL LEARNING: PARADOXICAL PREDICTIONS OF THE PAST

by

Jennifer Patricia Provyn

B.A., Rockhurst University, 2003
M.S., Syracuse University, 2007

Dissertation
Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Experimental Psychology.

Syracuse University
May 2013

Acknowledgements

This research was supported in part by a grant from the National Institute of Mental Health (MH-069938 to MWH). I thank Amy Criss, William Hoyer and Richard Aslin for their helpful comments on an earlier version of this document. I also thank Elfay Zhang and Asli Kilic for being dear labmates and colleagues as we all weathered the storms of graduate school.

At the core of pure research is an all-consuming and maddeningly addictive high. It is riding the wave of an idea, it is seeing a result no one has ever before seen. From formulating the idea, to the nuts and bolts of execution, to the critical transference of the results to the cultural knowledge base. It is thrusting staggered sheets of paper up to arcane flourescent lights, despite being surrounded by EEG rigs and Beowulf clusters, just to see if the error bars on those figures align, just to see if you have a result. It is taking code desconstruction so far into your head it cameos in your dreams. It is physical respite from the lab without ever having departed mentally. It is finally collapsing into sleep marathons. It is truly loving what you do. Sapere aude. Always.

I thank my advisor Marc Howard above all other mentors. In some sense, a pre-requisite for entry into the halls is noteable skill for learning facts and regurgitating information. But truly higher learning involves a development of ownership and mastery of a body of work, the ability to note patterns and devise tenable, testable solutions within that body, as well as the ability to generalize the skills gotten from practice in that niche, to the world at large. Marc taught me to formally, structurally, and defensibly think. He taught me to be my own devil's advocate. From Marc I also absorbed 4 other "facts about the world" that I contemplate regularly. 1. There's a bug in the code, the machine is only following commands. 2. Sells whats yous gots. 3. You've got to own it. 4. You don't have to take a class to learn. Just. Go. Figure it out.

Selfishly, in the pursuit of my goals, the lives of my family and friends have likewise in some way or another been forever altered. It can be a grueling and lonely gauntlet we chose for ourselves, with days that run into nights, and social obligations often collateral damage, an offering on the alter of higher aspirations. Just keep pushing. To the person who has always held my hand, whether I wanted it or not, who never lets me slide but who is the first to stand by my side: thank you, Mom, more than words can ever convey.

Contents

List of Figures

Introduction

In the natural world, some environmental elements change according to regular, periodic patterns over time. Prediction in these non-random environments can be guided by both the temporal co-occurrences of elements, as well as the probability that groupings, or sequences, of specific elements are followed by other elements. Organisms that can extract some degree of the statistical regularity present in the environment can in principle reduce uncertainty about future events to more accurately anticipate subsequent events. For example, it behooves a foraging animal to learn that the sight of moving prey often precedes the scent of the prey, and that the scent of the prey usually precedes its taste.

With respect to human learning, there is behavioral evidence indicating that the extraction of regularities from temporal sequences of events is a mechanism central to cognition. In addition to being central to cognition, this identification of patterns within sequences appears to occur very early in development. At two months of age, after only a few minutes of exposure to a series of visual stimuli that alternate between the left and right sides of a display, infants make anticipatory eye movements to the next event in the sequence (e.g., Canfield & Haith, 1991). By eight months of age, infants can identify word-like units in continuous speech after as little as 2 minutes of exposure (Saffran, Aslin, & Newport, 1996; Aslin, Saffran, & Newport, 1998). As demonstrated by infants, statistical learning of adjacent stimuli can occur without explicit awareness of the underlying structure of the input. Moreover, statistical learning has been shown to be domain general, with evidence for sensitivity to different stimulus probabilities across streams of continuous auditory input (Saffran, Johnson, Aslin, & Newport, 1999) and temporal sequences of visual shapes (Fiser & Aslin, 2002).

Serial Reaction Time Task

An overarching goal of the field of *statistical learning* is to provide a framework for studying prediction and inference. Central to statistical learning is the assumption that input, such as events in the environment, follow some unknown probability distribution. Following this assumption, successful prediction is typically dependent upon sensitivity to the distributional properties of the input environment. Evidence for having learned the distributional properties of the input environment is demonstrated by way of accurate predictions of the frequency and time-course of an events (Vapnik, 1998). Some methodologies for investigating sequential behavior, but which are not reviewed here, hail from the artificial grammar learning domain (Knowlton, Ramus, & Squire, 1992; Knowlton & Squire, 1994), as well as the procedural learning literature (for review see Willingham, Nissen, & Bullemer, 1989).

The serial reaction time (SRT) task, a model for studying sequential behavior, can be used to quantify sensitivity to sequential constraints present in structured environments (Nissen & Bullemer, 1987). During the SRT task, a fixed set of elements, such as letters, are presented to subjects (*S*s). Each element is associated in a one-to-one element-response mapping to a distinct and spatially unique motor response. Elements are presented individually and the *S*s task is to produce the motor response, such as pressing a specific button, that corresponds to each element. Unbeknownst to *S*s, there are predictive relationships among the elements. Evidence for sensitivity to the probabilities governing the transitions among elements is demonstrated through decreases in reaction time (RT) to predicted, relative to randomly presented, elements.

Consider the sequence A-B-C-A-B-C.... In this example, element A is a predictor for element B and element B serves as the prediction. Element B is a predictor for element C, with element C serving as the prediction. From the example, because element A always predicts element B, the probability that element A predicts element B is 1, $p(x_{i+1} = \text{B}|x_i = \text{A}) = 1$. A statistical learner's ability to extract this prediction statistic will be manifest behaviorally as faster RTs to the predicted element when it follows the predictor element, relative to RTs to the same element if that element is presented randomly. Returning to the example, given that element A is presented at time step i, RT to element B at time step $i + 1$ should decrease as the statistical learner becomes sensitive to the prediction statistics of the distribution.

In the classic version of the SRT task the time-course of statistical learning is assessed by comparing RTs between pseudorandomly assembled sequences of elements to sequences of elements with an embedded structure, usually much more subtle than the structure in the example sequence above. However, regardless of whether an SRT sequence of elements is pseudorandomly assembled or has structure, RT decreases universally across time for all of the elements as a function of practice and exposure to the SRT task. This universal decrease in RT is resultant from enhanced proficiency with the specific element-response mappings. Importantly, if structure is embedded into the sequence, then RTs to predictable elements decline below the baseline-RT of random-sequence elements. The decline in RT to structured sequence-elements is ascribed to sensitivity to relevant sequential contingencies, which presumably enable the learner to anticipate subsequent elements.

The SRT task can be implemented using either a deterministic or a probabilistic testing methodology. In the deterministic testing methodology, a fixed sequence of elements is repeated across blocks of determinstic trials, with transfer blocks of random trials inserted between blocks of deterministic trials (e.g, Nissen & Bullemer, 1987). Sequential learning is assesed by comparing performance on deterministic blocks with performance on random blocks. The structure of the deterministic and transfer block trials should be carefully controlled to ensure that sequential learning exhibited by Ss is not simply learning of relative frequency. Consider the Nissen and Bullemer (1987) ordered trial sequence: 4-2-3-1-3-2-4-3-2-1. Note that some positions occur more frequently than other positions (i.e., 1 and 4 occur 2x's; 2 and 3 occur 3x's). This sequence is in contrast to transfer blocks from the same study in which on average, each position occurred equally often. Since position frequency was not equated, RTs to deterministic trials could be faster than RTs to transfer blocks simply because Ss learned the nonsequential information that some positions were more likely to occur than other positions.

An alternative to deterministic sequence generation is probabilistic sequence generation. Probabilistic sequences can have noise randomly inserted into the sequence, or, more commonly, these sequences may be variants of deterministic sequences in which the conditional probabilities between elements has been maninpulated (e.g., Schvaneveldt & Gomez, 1998; Cleeremans & McCleeland, 1991). For example, the sequence fragment A B C may be followed by element D with probability 0.90, and by element E with probability 0.10.

There are three major limitations to the deterministic testing methodology that can be obviated with probabilistic sequence generation. First, in the deterministic sequence task design, learning is disrupted during the abrupt switch from deterministic to random blocks. In contrast, the probabilistic task design enables learning can be assessed continuously during training, without disruptions to Ss' representations of the task. Second, Ss can more easily explicitly learn sections of a sequence in the deterministic design, while learning is much slower with a probabilistic design. The attenuated learning rate associated with the probabilistic sampling allows for a larger number of observations to be collected (Cleeremans & McCleeland, 1991). Third, by its very design, a greater number of combinations of sequence elements can be represented in a probabilistic sequence than in a deterministic sequence. Thus, probabilistic sequence learning is more appropriate for examining Ss' sensitivity to statisical constraints, particularly higher-order associations.

Higher-order sequential learning involves developing sensitivity to relationships across structured input that extends beyond simple frequency effects and beyond simply learning strings of pairwise associations. In so far as tests of statistical learning are designed to measure sensitivity to higher-order associations, in addition to controlling for relative frequency, sequential trials must also control for pairwise associations between adjacent stimuli. Consider again the (Nissen & Bullemer, 1987) sequential trial sequence: 4-2-3-1-3-2-4-3-2-1. While the sequence does not contain uniquely predictive pairwise associations, the sequence does contain probabilistically predictive pairwise associations (Jackson & Jackson, 1992; Stadler, 1992). That is, 3 predicts 2 more often (2x's) than 3 predicts 1 (1x); 4 predicts both 2 and 3, but 4 never predicts 1. *S*s who are sensitive to the probabilistically predictive pairwise associations may exhibit faster RTs for sequential trials than for pseudorandom trials, without having necessarily learned higher-order relationships beyond these first-order conditional associations. First-order conditional sequence learning will first be discussed, followed by higher-order sequence learning.

First-Order Conditional Sequences

The simplest kind of sequence of events is a Markov chain, which is a sequence composed entirely of adjacent predictive elements in which each sequence element at time-step i, (x_i), unambiguously predicts the next sequence element (x_{i+1}). In the statistical learning literature, these Markov chains are referred to as deterministic first-order conditional (FOC) sequences. The probability of a transition between consecutive elements in a FOC sequence is given by the conditional probability $p(x_{i+1}|x_i)$. For example, given the sequence, A-B-C-A-B-C..., $p(x_{i+1} = B|x_i = A) = 1$.

While the simplest type of sequence learning involves fixed sequences most sequences of events in the natural world contain a mix of patterned and un-patterned, or random, variability. The computational problem of finding structure in a continuous stream of experience that can then be used to guide prediction is multi-tiered: reusable units are first detected; patterns are then generalized or inferred across the units; finally, the predictive value of the patterns across the units can be assessed to enable accurate prediction of future events. As an example consider human infant learning. Much of human infant learning is organized serially (over time), including locomotion, social interaction and ultimately, language (Goldstein et al., 2010; Lashley, 1951). Elements can be detected from a stream of continuous input and later predicted by computing the likelihood that element X predicts element Y.

More similar to the natural world in which events are not perfectly predictable, probabilistic sequences are composed of relationships between elements in which at least some of the elements predict other elements with probability < 1. Consider a probabilistic FOC sequence in which element X is a predictor for both elements Y and Z with unequal probabilities. Specifically, assume that element X predicts element Y on 80% of the trials in which it is presented and element Z on the remaining 20% of the trials in which it is presented. As Ss become sensitive to the probabilistic prediction statistics for element X, the conditional uncertainty for elements presented immediately after X should decrease. Given that element X is presented at time step i, if a sequence learner has extracted the prediction statistics of element X, then his/her RT to the element presented at time-step $i+1$ should be faster if that element is Y, than if that element is Z.

Transition probabilities (TPs) are prediction statistics that quantify the strength with which x predicts y. TPs are calculated according to the equation $TP = p(y|x) = \frac{frequency(xy)}{frequency(x)}$. This conditionalized statistic tracks the frequency that elements co-occur in a particular order, normalized as a function the element's overall frequency in the corpus. TPs are important for extracting temporally co-occurring sequences of phonemes from continuous input, such as identifying the boundaries between words in fluid speech.

Adjacent Statistical Segmentation

Adjacent Linguistic Statistical Segmentation

The continuous nature of speech makes word segmentation a particularly challenging task. Fluid speech is not characterized by words delineated by obvious acoustic cues, such as pauses between word boundaries (Cole & Jakimik, 1980; Saffran, 2003). For example, infants are not innately equipped with the knowledge that *pretty* and *baby* are words, while the sequence of letters *tyba* spanning the words' boundaries is not a word (Saffran et al., 1996).

To successfully segment words from a continuous stream of sounds, Ss must first discriminate the speech syllables, then track the temporal order in which these syllables occur, and finally, track the probability of these orderings (Newport, Hauser, Spaepen, & Aslin, 2004). Despite the difficulty of word segmentation, infants as young as 7.5-months of age are able to extract words from continuous speech streams by tracking the co-occurrence statistics of adjacent elements (Jusczyk & Aslin, 1995). Saffran et al. (1996) exposed 8-month olds, first graders and adults to continuous sequences of spoken nonsense languages composed of multisyllabic words (e.g., *golabupabikututibubabupugolabubabubu...*). Ss were tested on whether they could discriminate words from the language (e.g., *golabu*) with sequences that spanned word boundaries (e.g., *bupabi*). Results confirmed that all three groups of Ss could discriminate word boundaries, which the authors suggest is evidence for sensitivity to the statistical properties of the languages.

When tracked across a corpus of sounds, the TP between two sounds is typically higher within words than between words (e.g. Harris, 1955; Saffran et al., 1996; Saffran, Newport, & Aslin, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; Aslin et al., 1998). This ability to naturally and automatically encode statistical regularities in speech streams without overt guidance or reward appears to begin even before infancy, while in utero, with newborns demonstrating preference for speech in the mother's language as compared to other languages (Moon, Panneton-Cooper, & Fifer, 1993).

Adjacent Non-Linguistic Statistical Segmentation

In addition to evidence that statistical segmentation is present in linguistic domains, behavioral studies have shown that humans can extract statistical regularity in scenes of visually displayed shapes. Each visual scene is composed of abstract shape elements. The shape elements are concatenated into visual chunks in which two or more spatially adjacent shape elements always co-occur in the same relative configuration. There are no obvious cues to segment the identity of the visual chunks other than co-occurrence statistics. After familiarization with the visual scenes, *S*s reliably choose fragments of visual chunks over random combinations of shapes. Moreover, both adults (Fiser & Aslin, 2001) and infants (Kirkham, Slemmer, & Johnson, 2002) are able to detect the statistical consistencies among adjacent shapes to group them into shape "words".

In addition to static visual scenes of shape "words", Kirkham et al. (2002) habituated young infants to sequences of predictable discrete visual stimuli. Infants were then tested on the statistically predictable sequence alternating with a novel sequence of identical discrete visual stimuli. Infants exhibited significantly greater interest in the novel sequences. Fiser and Aslin (2002) likewise presented adults with a continuous temporal sequence of shapes. Despite the fact that the frequency of individual shapes across the sequence was equated, adults demonstrated sensitivity to the temporal structure of the sequence.

These results suggest that statistical learning of adjacent dependencies extends across development and domain, with infants and adults alike able to extract the statistical information for shape word boundaries with potentially the same mechanisms as auditory word boundaries. Further support for the hypothesis that similar mechanisms for extracting statistical regularity operate across domains is present in studies of language impairment. Children with specific language impairments have corresponding difficulty with non-language, visual sequence learning (Tomblin, Mainela-Arnold, & Zhang, 2007). Some authors suggest that the ability to learn from experience through statistical learning contributes to, if not mediates, subsequent linguistic performance (Misyak, Christiansen, & Tomblin, 2010).

N^{th}-Order Conditional Sequences

Prediction tasks often involve higher-order non-adjacent dependencies in which recursive pairwise associations, such as those present in FOC adjacent dependencies, are non-predictive. Instances of non-adjacent dependencies in the natural environment are abundant in language comprehension and production. In English, for example, auxiliaries and inflectional morphemes (e.g., am typing, has worked) as well as number agreement dependencies (the dogs in the yard are dirty) are separated by intervening elements.

Recurrent connectionist models have been successfully applied to SRT learning (e.g., Cleeremans & McCleeland, 1991; Cleeremans, 1993) and in general, a mechanism that is sensitive exclusively to FOC associations between adjacent stimuli is computationally insufficient to model higher-order sequence learning. Prediction of higher-order non-adjacent dependencies is contingent upon some combination of preceding elements (Reed & Johnson, 1994; Reber & Squire, 1994; Curran, 1997; Schendan, Searl, Melrose, & Stern, 2003). Consider element B from the repeating second-order conditional (SOC) sequence A-B-A-D-B-C-D-A-C-B-D-C. B appears in three different FOC prediction contexts, B-A, B-C and B-D, so is not uniquely predictive of any one element. Computationally, however, the representation of the first B is different from the representations of the second and third B's because in each instance, element B is preceded by different elements (i.e., A-B, D-B and C-B). For accurate prediction, learners must extract both the FOC associations between immediately adjacent elements *and* extend the temporal context back two time steps to generate a unique prediction. For example, if $x_i = B$ and $x_{i+1} = C$, then

$x_{i+2} = D$, or $p(x_{i+2} = D | x_i = B, x_{i+1} = C) = 1$. Insofar as the representation of each element is a function of all previous elements, higher-order associations (e.g., third-, fourth-order) can develop. Sensitivity to these higher-order, non-adjacent dependencies emerges more gradually than sensitivity to FOC, adjacent dependencies (Cleeremans & McCleeland, 1991).

Non-Adjacent Statistical Segmentation

Consider again the previous examples of instances of non-adjacent dependencies in the English language: auxiliaries and inflectional morphemes (e.g., am typing, has worked), and number agreement dependencies (e.g., the dogs in the yard are dirty). Particularly in the example of number agreement dependencies, the elements of agreement, dogs and are, are separated by the irrelevant intervening elements, "in the yard". Non-adjacent dependencies characterized by irrelevant intervening input can be examined with artificial strings of the form aXd and bXe, where the relations are a_d and b_e, and X is completely non-predictive. Infants and adults display greater sensitivity to the non-adjacencies when X is drawn from a large pool of elements as compared to a small pool of elements (Gomez, 2002). In other words, when the context for the intervening elements is either not variable and therefore relevant to prediction (e.g., X is drawn from a pool of 1) or context is highly variable and therefore irrelevant to prediction (e.g., X is drawn from a pool of 18 or 24), Ss are best able to detect invariant structure (Onnis, Christiansen, Chater, & Gomez, 2003; Gomez, 2002).

Evidence that statistical learners are sensitive to non-adjacent dependencies appears in auditorily presented material from linguistic- (Gomez, 2002; Perruchet, Tyler, Galland, & Peereman, 2004; Onnis, Monaghan, Christiansen, & Chater, 2004; Newport & Aslin, 2004; Aslin et al., 1998; Saffran et al., 1996, 1997) and non-linguistic domains (i.e., tone sequences) (Kuhn & Dienes, 2005, 2008; Creel, Newport, & Aslin, 2004; Saffran et al., 1999). Additionally, non-adjacent contingencies derived from statistically structured material have been observed in visually presented shape arrays (e.g. Fiser & Aslin, 2001, 2002; Kirkham et al., 2002). The structural relationships in non-linguistic statistical learning is subject to spatial constraints analogous to the temporal constraints present in linguistic statistical learning (Conway, Goldstone, & Christiansen, 2007).

Although human adults and infants readily extract regularity among both immediately adjacent and non-adjacent elements, there are limitations to the groups' sensitivities to temporal order in studies of non-adjacent sequence learning. These limitations to non-adjacent sequence learning are evident in language learning, with manipulation of the non-adjacency between syllables, consonants and vowels.

Non-Adjacent Syllables. Newport and Aslin (2004) presented subjects with continuous streams of speech in which patterned relations among syllables occurred between non-adjacent syllables. The non-adjacent syllables were separated by an intervening unrelated syllable. Results confirmed that human adults are not readily able to acquire an artificial language in which words of the language are composed of regularities among non-adjacent syllables. The authors note that natural human languages also do not contain words formed from a stem consisting of related syllables 1 and 3.

Non-Adjacent Phonemic Segments (Consonants). While human languages do not frequently contain word-formation patterns consisting of non-adjacent syllables, a common non-adjacency pattern in human languages are word-formations consisting of non-adjacent phonemic segments (consonants) (Newport & Aslin, 2004). For example, Semitic languages such as Hebrew form many words out of a three consonant stem (i.e., *k-t-b*, meaning "to write") (Newport et al., 2004). Vowels inserted between the consonants vary contingent upon the tense of the word. Learners must therefore attend to consistent patterns among consonants. Newport and Aslin (2004) tested adult English speakers on streams of continuous speech in which words of the language were composed of regularities of non-adjacent consonants. Results confirmed that Ss were able to acquire the regularity (Newport & Aslin, 2004).

Non-Adjacent Phonemic Segments (Vowels). Non-adjacent vowel segments are another common non-adjacency word-formation pattern. Consider, for example, Turkish "vowel harmony". Vowel harmony occurs when vowels spanning a word agree with one another in certain features, like place of articulation (Newport et al., 2004). Learners of the Turkish language must attend to consistent patterns among vowels because the consonants inserted between the vowels vary. Although the ability to monitor word-formation patterns composed of non-adjacent vowel segments is critical for some languages, the development of this non-adjacency tracking does not appear to be contingent upon exposure to these languages. For example, the English language does not contain vowel harmony, none the less, when tested on streams of continuous speech in which words of the language were composed of regularities of non-adjacent vowels, adult English speakers were able to acquire the regularity (Newport & Aslin, 2004).

In summary, statistical learning is not limited to elementary computations on immediately adjacent syllables, with adult learners demonstrating selective types of non-adjacent statistical learning. The lack of uniformity across linguistic non-adjacency learning might be mediated by processing mechanisms. Newport and Aslin (2004) suggest that Ss may process dependencies in terms of element-level segments, at the level of individual consonants and vowels, which would complicate tracking non-adjacent syllable regularities. The authors further suggest that the selective sensitivity to individual element-level non-adjacent dependencies, coupled with the lack of sensitivity to syllable-level non-adjacent dependencies, may have helped to shape human languages.

Representations of Sequences

Work from the adjacent and non-adjacent sequence learning literature provides evidence that statistical learning is a domain-general, fundamental mechanism that contributes to the development of internal representations of the environment. Transfer learning tasks can be employed to tease apart these internal representations. That is, whether the internal representations are of the stimulus, the response, or some intermediate representations therein. To isolate the components of the internal representations, Cohen, Ivry, and Keele (1990) kept the stimuli constant, but modified the response sequence to the stimuli in a sequential learning task. Despite the change in response representations, the authors observed transfer of learning. These results would seem to support a stimulus-based representation account of sequence learning since modifying the response representation did not detrimentally affect sequence learning gains. To test the limits of stimulus-based representations, Keele, Jennings, Jones, Caulton, and Cohen (1995) manipulated the response-sequence modality from manual to verbal. This extreme change in response-representation resulted in incomplete transfer. Because some of the learning gains were lost in the transition from manual to verbal responses, the authors suggested that sequence learning is not entirely stimulus based.

To examine if the internal representations of the environment formed during sequence learning are response-based, Willingham (1999) initially instructed Ss to respond to spatial locations of stimuli using an incompatible response-key mapping. The stimulus-sequence was then changed such that the the response-key mapping was compatible with the stimulus mapping. Results confirmed that the sequence learning transferred to the new condition, as long as the order of the response-key presses remained the same as in the initial learning condition.

With data from the transfer learning literature supporting both stimulus- and response-based representations of the environment, it is likely that what is being represented in sequence learning is some combination of both the stimulus and response contexts. This hypothesis is consistent with within-domain evidence that the products of statistical learning are fairly abstract and generalizable. In the visual domain, Ss presented with colored visual stimuli at familiarization were able to abstract these regularities to black shapes during test (Turk-Browne, Junge, & Scholl, 2005). In the auditory domain, Ss were able to generalize from non-distorted input at familiarization to distorted input at test (Vouloumanos, Brosseau-Liard, Balaban, & Hager, 2012). While segmented units are fairly abstract and generalizable, the products of statistical learning do not transfer particularly well across modalities (e.g., from auditory to visual stimuli) (Conway & Christiansen, 2006).

Explicit Knowledge

An implicit learning task is one performed without the S's awareness of, or conscious effort to use, memory representations to influence performance with items that had been previously presented. Sequence knowledge is assumed to be implicit insofar as Ss demonstrate sensitivity to the underlying sequence structure in the context of an indirect test of learning. If the sequential representation is in fact conscious, then Ss should employ this knowledge when instructed to so do in an explicit test of knowledge (Merikle & Reingold, 1991).

The most standard recall tests of explicit knowledge range from free verbal reports to unstructured questionnaires (e.g. Willingham et al., 1989; Lewicki, Hill, & Bizot, 1988). For example, Perruchet and Amorim (1992) employed a "free" generation task in which Ss were instructed to generate complete sequences of trials, absent feedback. Other authors have proposed recognition tasks in lieu of recall, wherein Ss are presented with sections of sequences and instructed to judge the likelihood that the sequence appeared during the SRT task (Willingham, Greenley, & Bardone, 1993; Perruchet & Amorim, 1992). However, both verbal report- and recognition paradigms have been characterized as weak methods for assessing explicit knowledge of implicit learning (Jackson & Jackson, 1995; Perruchet & Amorim, 1992; Shanks & St. John, 1994) because the test context is incongruent with the learning context of the SRT task.

Nissen and Bullemer (1987) developed an explicit test of sequence learning specifically designed to be more contextually equivalent to the SRT task. This "standard" generation test was a cued-recall design in which Ss were presented with a visual stimulus from the previously performed SRT task, and were instructed to explicitly predict the stimulus that should come next in the sequence. On each trial, the stimulus appeared below one of six screen positions and Ss had to press the key corresponding to the position at which they expected the next stimulus to appear. The standard generation task differed from the SRT task in that Ss were instructed to *generate* a response and to respond *slowly*. Accuracy, rather than RT, was the primary dependent measure. Ss were instructed to keep guessing until a correct prediction was produced–such that several guesses could occur between any two trials of the task–at which point the next stimulus was presented and the next prediction trial initiated. Accurate performance on the standard generation task was taken to reflect explicit knowledge of the SRT sequence. Despite demonstrating sensitivity to the relationships among the sequence elements in the SRT task, Ss were not able to accurately generate these relationships in the explicit prediction task (Nissen & Bullemer, 1987; Cohen et al., 1990; Willingham et al., 1989).

While more contextually similar to the SRT task than recognition or explicit recall tasks, the standard generation task and the SRT task are arguably still procedurally distant. Moreover, the multiple guessing design of the generation task may induce memory interference such that the responses produced by Ss on each trial could interfere with memory of previous elements (Perruchet & Amorim, 1992; Jimenez, Mendez, & Cleeremans, 1996). An alternative to the standard generation task, the "continuous" generation task is a more contextually similar, direct test of SRT learning (Cleeremans & McCleeland, 1991; Cohen et al., 1990; Shanks & Johnstone, 1999; Jimenez et al., 1996).

During the continuous generation task, Ss are required to predict the next stimulus on each trial in a stimulus-prediction-stimulus-prediction design. Instructions emphasize accuracy, rather than RT, and consistent with the SRT task, Ss' responses are captured via keypress rather than via recall. Some designs (e.g., Cleeremans & McCleeland, 1991) sound a beep for incorrect predictions. In contrast to the standard generation task, the next stimulus presented on each trial is defined by the sequential structure, regardless of Ss' prediction responses.

The role of awareness, or explicit knowledge, in a sequential learning task may not be a necessary condition for statistical learning. Instead, the interaction of sequential learning system(s) with other neural areas could cause the emergence of explicit knowledge. While explicit knowledge is not a prerequisite for successful performance on the SRT task, it has been shown to enhance SRT learning (Perruchet & Amorim, 1992; Willingham et al., 1989). For example, Ss who acquire explicit knowledge of the underlying sequence structure demonstrate more anticipatory responding and faster RTs (Curran & Keele, 1993; Willingham et al., 1989) than do Ss with little to no explicit knowledge. Explicit knowledge may enable Ss to generate the next stimulus prior to stimulus onset, while implicit knowledge may reflect a priming process that facilitates responses but does not enable explicit recall. Curran and Keele (1993) found that Ss who explicitly learned a sequence demonstrated a RT advantage in comparison to Ss who implicitly learned the same sequence. However, this enhanced SRT performance mediated by explicit knowledge disappears if Ss are transferred to a distraction condition (Curran & Keele, 1993).

A variety of factors can influence Ss' ability to obtain explicit knowledge of the relationships between sequence elements. Ss can be explicitly informed of the presence of structure in the sequence of stimuli. The complexity of the sequence, that is, whether it is deterministic or probabilistic, will affect whether Ss become explicitly aware of the structure. The presence or absence of a distractor can attenuate or altogether eliminate the ability to acquire explicit learning. In addition to explicit instructions, the complexity of the sequence structure, and the presence or absence of a distractor, the response-stimulus interval (RSI) can also influence the acquisition of sequence knowledge. Frensch and Miner (1994) found implicit learning when RSIs were brief (500ms), while explicit learning was significant after longer intervals (1500ms). The authors suggest that in the implicit sequence learning condition, the stimuli become associated together by being co-activated in short-term memory. However, this co-activation in short-term memory only persists for short RSIs. In contrast, the authors suggest that explicit SRT learning is related to a working memory contribution (particularly given the sensitivity to distraction found in explicit learning conditions). Longer RSIs would enable active rehearsal processes to maintain the activation of the stimuli across longer time periods, which could result in explicit knowledge of the relationships between sequence elements.

Episodic Memory

Recall that both adjacent and non-adjacent dependency statistical learning has been demonstrated across human development and stimulus modality. It is therefore certainly conceivable that learning mechanisms not necessarily "designed" for a specific application may mediate both prediction, and more generally, statistical learning. In addition to tasks historically classified as tests of "statistical learning" (e.g., the SRT task), sensitivity to statistical regularity in the environment appears to be a critical dimension of various episodic learning methodologies.

While statistical learning is typically understood as developing the ability to predict the future, episodic memory is typically understood as the ability to remember specific instances from the past. Although seemingly disparate paradigms, consider the similarity between the episodic memory paired-associate learning task and the statistical learning deterministic FOC sequence learning task. In paired-associate learning (e.g., A-B, C-D), the first element of each pair (e.g., A, C) serves as a cue for the recall of the second element (e.g., B, D). The literal responses required in each task differ (i.e., recall vs. motor responses). However, within each methodology the task of the learner is to make a future response based upon the prediction generated by the probe element. The relationship between elements of each pair in paired-associate learning is similar to that of the relationship between two sequential elements in a deterministic FOC sequence.

It has been extensively documented in paired-associate learning, as well as across a variety of episodic recall paradigms, that graded associations can be formed among elements co-occurring in close temporal proximity (e.g., the first and second elements of a pair sequence; for a review, see Kahana, Howard, & Polyn, 2008). As such, it has been hypothesized that temporal contiguity could be the essence of, or at the very least an underlying mechanism facilitating, episodic memory (Sederberg, Howard, & Kahana, 2008). Kahana (1996) introduced the lag-recency effect to illustrate the associated structure of item learning in episodic memory. With respect to free recall, after having learned a list of words, successively recalled items have a higher probability of originating from nearby serial positions than remote serial positions. This lag-recency effect is measured by the conditional response probability curve (CRP curve, Howard & Kahana, 1999; Kahana, 1996). The CRP curve is plotted as a function of the lag, or distance in the list between studied items. The CRP curve is characteristically peaked in the middle around lag zero, indicating recalls are more probable around nearby serial positions. It is also asymmetric, with higher conditional probabilities of recalling an item in the forward than in the backward direction.

In serial list learning, elements can be associated together by virtue of their temporal order. For example, having recalled element "A" increases the probability that element "C" will be recalled over element "F". While a temporal contiguity mechanism may be sufficient for associating serial elements together, the formation of associations among non-adjacent elements through temporal contiguity could place an unmanageable computational burden upon learners. In other words, if elements must co-occur close

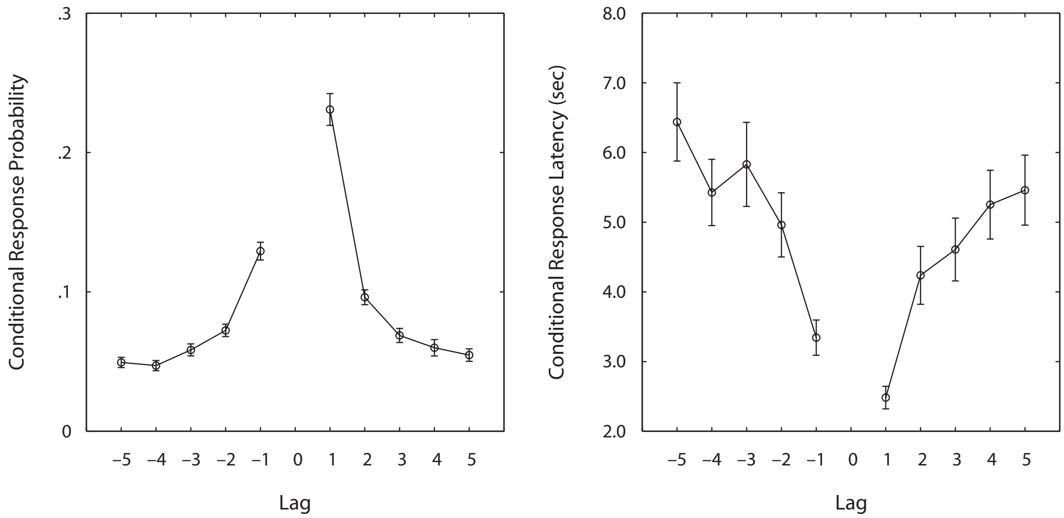

Figure 1: Conditional response probability (CRP; left) and conditional response latency (CRL; right) as a function of serial position lag. Error bars represent 95% confidence intervals. **Figure originally presented in (Sederberg et al., 2010)**.

together in time to become associated together, then the predictive relationship between elements not presented together in time could quickly become obfuscated as a function of the quantity of intervening elements.

Kilic, Criss, and Howard (2013) demonstrated both short- and long-term contiguity effects in a probed recall task in which test order was specifically disrupted to be uncorrelated by the use of multiple study lists. A short-term contiguity effect was evidenced by the data that Ss tended to generate words from nearby serial positions if the generated words were from the same list as the probe item. Importantly, there was simultaneous evidence for long-term contiguity effects in the experiment: given that Ss recalled a word from a different list from the probe item, words tended to come from nearby lists. The Kilic et al. (2013) results provide evidence against temporal contiguity, as well as correlations between study and tests contexts, as the mechanisms exclusively driving the associations present in these contiguity effects.

Howard et al. (2009) examined the associative structure induced by learning double-function lists of paired-associates (e.g., A-B, B-C) and demonstrated graded contiguity effects. The graded contiguity effects reflected the ordered-sequence from which elements had been drawn, and importantly, represented associations formed among words that were never presented together in time (see also, Popper, 1959; Slamecka, 1976; Bunsey & Eichenbaum, 1996). Rather than reflecting a gradient of temporal contiguity, Howard et al. (2009) suggested that the associations formed among elements that did not temporally co-occur reflected a gradient of the underlying structural relationships among the study elements.

The graded associations formed among elements in the Howard et al. (2009) study extended in both the forward and the backward directions. Backward associations are a ubiquitous result across a variety of episodic recall tasks (e.g. Primoff, 1938; Kahana et al., 2008; Slamecka, 1976; Kahana, 1996; Klein, Addis, & Kahana, 2005) and are beneficial to the extent that they increase the probability of recalling sequence elements. However, unlike free and serial recall tasks, backward associations, and graded associations in general, are anti-predictive in paired-associate learning tasks. That is, there is no benefit to predicting an element other than the one being probed. In fact, backward associations are the primary source of interference in paired-associate learning tasks (Primoff, 1938; Umemoto & Hilgard, 1961; Young, 1961; Slamecka, 1976; Howard et al., 2009; Provyn, Sliwinski, & Howard, 2007). The formation of associations beyond the immediately predictive element in prediction tasks is suggestive of a ubiquitous underlying associative mechanism, which universally gives rise to graded contiguity effects, regardless of the specific application (Howard et al., 2009).

While Howard et al. (2009) provided evidence in an episodic learning task, Hunt and Aslin (2010) provided promising evidence for transitive associations among elements in a non-episodic learning task. The authors constructed “grammars” of elements grouped into categories. Using an SRT task, they constrained the distributional information available from the presentation sequence. The extraction of the categorical membership of the stimulus elements was dependent upon the stimulus elements’ prior and posterior probabilities during training and test, rather than on the elements' serial order positions. Results demonstrated that Ss were able to induce categories on the basis of distributional information, but also that they became differentially sensitive to variations in the transitional statistics that defined the categories.

Sensitivity to Backward and Graded Associations

In addition to forward transitional probabilities, Aslin et al. (1998) suggested that backward associations, while perhaps not informative in standard SRT prediction tasks, are useful for discovering some relationships in language learning. For example, backward transitional probabilities are far more informative than forward transitional probabilities for discovering the grammatical category "noun" (Willits, Seidenberg, & Saffran, 2009). The specific degree of sensitivity to backward and forward probabilities may in fact be mediated by the learner's natural language. For example, in languages with grammatical gender, sensitivity to backward statistics should facilitate acquisition of linguistic structure (Pelucchi, Hay, & Saffran, 2009).

Studies that manipulate forward transitional probabilities typically do not control for backward transitional probabilities. Both probability statistics are typically correlated, complicating the task of teasing apart independent roles. Pelucchi et al. (2009) specifically measured whether 8-month-old infants track backward transitional probabilities in continuous speech of disyllabic sequences. Test words occurred equally often during familiarization, shared the same trochaic stress pattern and, importantly, had the same forward transition probability ($p = 1.0$). After familiarization to the speech stream, infants were tested on high transitional probability words (where $backward\ transition\ probability = 1.0$) and low transitional probability words (where $backward\ transition\ probability = 0.33$). Infants were tested using the Head Turn Preference Procedure (Saffran et al., 1996). During the familiarization phase, infants listened to the language projected from speakers mounted beneath two lights. The lights served to maintain infant attention and flashed contingent upon looking behavior while the familiarization sequences played continuously. After familiarization, infants were tested on trials of single items. Each test item was played and repeated as long as the infant maintained a head-turn in the direction of the flashing light above the speaker projecting the sound. Infants demonstrated sensitivity to the backward transitional probability statistics with significantly longer looking times, on average, to high backward transitional probability words than to low backward transitional probability words. Perruchet and Desaulty (2008) likewise demonstrated that adults use both forward and backward TPs for word segmentation.

Examining Backward and Graded Associations

The following experiment employed a probabilistic SRT task to quantify sensitivity to backward and remote associations. Stimuli consisted of 11 randomly-selected letters arranged into an underlying ordered-structure (Figure 2A). Letters were probabilistically sampled from the structure and assembled into sequences such that the stimulus presentation traversed the structure in a clockwise direction within the structure. The transition probability to the sequentially forward-adjacent element was .7. For example, from Figure 2A, given that L had just been presented, the probability of W next being presented was $p(.7)$. These transitions were referred to as lag_{+1}. All other letters from the range of 10 possible letters, excluding the current letter itself—the same element was never presented in succession—were sampled probabilistically from the ring with probability $p(.3/9)$. These elements were collectively referred to as lag_{RM}.

The goal of the experiment was to examine whether there would be quantitative performance differences to lag_{RM} element transitions drawn from an ordered-structure in which each lag_{RM} transition was constrained to be probabilistically equivalent. The lag_{RM} associations counter clockwise in the ordered-structure from lag_{+1} actually reflect a binding of an element to elements that preceded it in the past and, as such, are arguably the opposite of prediction. Interpreted at the task-level, in so far as the goal is to predict the next clockwise element in the sequence, consistent counter clockwise predictions are paradoxical. Consistent counter clockwise predictions represent predictions of the past, which are uninformative and incorrect responses in the context of the SRT task. Beyond the SRT task, however, a member of the lag_{RM} category, lag_{-1} (e.g., W $\rightarrow$ L) associations have been demonstrated across a variety of recall tasks, and sensitivity to lag_{-1} statistics are suggested to facilitate acquisition of linguistic structure (Pelucchi et al., 2009).

Based upon the probabilistic learning environment of this experiment, because the probability of lag_{-1} was the same as the probability of any other lag_{RM}, lag_{-1} should not be better predicted than any other lag_{RM}. That is, RTs to the transition $i \rightarrow i - 1$, should not differ from the RTs to $i \rightarrow i \pm x$, where x is any lag from 2 to 5. Sensitivity to backward associations in this experimental context would be manifest as a reduction in RT to lag_{-1} predictions. For example, given that $x_i = \text{L}$, evidence for a backward association would be faster RT to $x_{i+1} = \text{V}$, relative to other lag_{RM} transitions. In addition to sensitivity to lag_{-1} associations, lag_{RM} RTs were predicted increase as a function of the absolute value of lag. For example, RT to the transition $p(x_{i+2}|x_i)$ was predicted to be faster than RT to the transition $p(x_{i+5}|x_i)$.

Experiment

The experiment specifically examined the question of whether backward and graded associations can be observed in a statistical learning paradigm. Letter-elements, henceforth "elements", were individually presented in runs of ordered test-sequences in an SRT task. Elements from the test-sequences were visually presented on a computer screen. The subjects' task was to respond, via key-press, to each element. Test-sequences were systematically assembled from an underlying ordered-structure. To create the test-sequences, at each time-step i, the probability of moving forward one slot in the ordered-structure was .7. The probability of jumping to any other element in the ordered-structure was .3/9.

FOCs were probabilistically constrained to examine whether subjects would demonstrate sensitivity to the underlying ordered-structure from which the elements were sampled. If subjects were sensitive only to the probability of each lag transition, then at each time-step i, RTs should be fastest to elements presented one slot forward in the ordered structure, and RTs should not differ significantly across the other element transitions. However, if subjects were also sensitive to the underlying ordered-structure from which the elements were sampled, then RTs should differ as a function of the distance within the ordered-structure between the element presented at time-step i and the element presented at time-step $i + 1$. A lag statistic (e.g., Kahana, 1996; Howard et al., 2009) was used to quantify associations within the ordered-structure.

Lag is defined as the distance between two elements in an ordered-structure. Larger absolute values of lag indicate greater distance between the elements, while the sign of the lag indicates the direction of the distance. A positive lag represents a transition to an element in the clockwise direction within the structure, while a negative lag represents a transition to an element in the counter-clockwise direction. A function of the circular nature of the ordered-structure, each transition could be defined by two lags. For example, from Figure 2A, Z $\rightarrow$ V is both a lag$_{+2}$ and a lag$_{-9}$ transition. Unique lag values therefore were constrained to the ranges $[-5, -1]$ and $[1, 5]$. Lag$_0$ was excluded from the range of legal lags because the same element was never presented in succession.

The relationship between the element at time-step i and the element clockwise one slot in the ordered-structure, is defined as lag$_{+1}$. Illustrated in Figure 2A and B, an example of a lag$_{+1}$ transition is Z $\rightarrow$ O. All other lags were referred to as lag$_{RM}$. The lag$_{RM}$ category was further delineated into three specific groups: lag$_{-1}$; lag$_{+RM}$; and lag$_{-RM}$. A lag$_{-1}$, or backward transition, was defined as a transition from the element at time-step i to the element one slot counter-clockwise in the ordered-structure. From Figure 2A and B, an example of a lag$_{-1}$ is the transition L $\rightarrow$ V. A lag$_{+RM}$ transition was defined as a non-lag$_{+1}$ transition in the clockwise direction in the ordered-structure from the element presented at time-step i. An example of a lag$_{+RM}$ transition from Figure 2A and B is L $\rightarrow$ P. Due to the circular nature of the ordered-structure, lag$_{+RM}$ values could range from lag$_{+2}$ to lag$_{+10}$. Note that a lag$_{+10}$ transition, Z $\rightarrow$ Q, has already been defined as a lag$_{-1}$ transition. From the specific hypotheses motivating the experiment, lag$_{+RM}$ values were constrained to the range of clockwise transitions from lag$_{+2}$ to lag$_{+5}$.

A lag$_{-RM}$ transition was defined as a counter-clockwise transition in the ordered-structure from the element presented at time-step i. Lag$_{-RM}$ values, again due to the circular nature of the ordered-structure from which the test-sequences were generated, could range from lag$_{-2}$ to lag$_{-10}$. Again from the specific hypotheses motivating the experiment, lag$_{-RM}$ values were constrained to the range of counter-clockwise transitions from lag$_{-2}$ to lag$_{-5}$.

This experiment provides two distinct and unique contributions to the literature. First, the exeperimental design marries the SRT methodology for measuring sensitivity to statistical regularity in structured environments with the analytic techniques limited heretofore to episodic memory methodologies. The experiment therefore serves as a bridge between the two literatures. Second, the simplicity of the underlying ring structure from whence stimuli were probabilistically sampled lends itself precisely to examining sensitivity in the learning environment, without having to extract variance or otherwise control for complex finte-state grammatical structures.

To quantify explicit learning of the ordered-structure in this experiment, an "interposed generation" set of trials followed the final SRT trial. Element sampling and presentation in the interposed generation trials were identical to that of the SRT trials, with the exception that occasionally *S*s were required to explicitly predict which element should come next in the sequence. If *S*s are sensitive to the FOC probabilities–wherein each element predicted another element with probability .7–then predictions should be primarily composed of lag_{+1} transitions. Additionally, if *S*s had become sensitive to the ordered-structure from which elements were probabilistically sampled, then errant predictions should be composed of more predictions to promixal lags, with the probability of errant predictions decreasing as a function of the absolute value of lag.

Importantly, there are two reasons that the interposed generation task is a unqiue and distinct contribution to the literature as a method for quantifying explicit SRT learning. First, in both the standard and continuous generation tasks, extended practice and exposure to the task offers a new learning opportunity for *S*s. The traditional way to address this issue has been to limit generation trials, with some authors suggesting that only the first few repetitions of a sequence can be considered relevant for analysis (Willingham et al., 1989). The interposed generation task largely removes the task as a learning opportunity. Recall that the continous generation task requires the *S* to produce a prediction at every trial in an element-predict-element-predict format, with some designs even incorporating feedback for incorrect responses. In this experiment, since all lag_{RM} elements were equally likely to be presented, prediction of a lag_{RM} element is not incorrect, nor does it necessarily lead to new learning by virtue of the fact that the generation trials were interposed into the stream of the SRT task, rather than in an element-predict-element-predict format.

Second, both the standard and the continous generation tasks (though see (Jackson & Jackson, 1995)) emphasize accuracy. Statistical knowledge in a probabilistic learning environment, however, is reflected in sensitivity to different transition probabilities between sequentially adjacent elements. The ring structure in this experiment was traversed probabilistically such that the most statistically likely transition, lag_{+1}, was considered "accurate" from a serial-order perspective. Beyond lag_{+1} generated-responses, because each lag_{RM} transition in the SRT task was equally likely, the pattern of lag_{RM} generated-responses provides insight into Ss ability to express explicit awareness of the underlying ring structure from which elements were probabistically sampled.

Methods

Subjects

A power analysis using GPower (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that for an 80% chance of detecting a medium effect (defined by Cohen (1992) as .5 of a population standard deviation between the means) at the Bonferroni corrected .005 (one-tailed) level, the mininum total sample size would be 192. Two-hundred and seventeen undergraduate students participated for course credit in an introductory psychology course at Syracuse University. Twelve subjects were eliminated from analysis due to errant responses and/or spurious key-presses comprising $\geq 10\%$ of their total number of responses.

Materials

Stimulus-elements consisted of the complete 26-letter English alphabet. Lists of elements were formed for each subject by first randomly sampling 11 elements without replacement from the alphabet. Each element was then assigned to a slot in an ordered-structure (Figure 2A). Test-sequences were generated from the elements in the ordered-structure.

Each test-sequence was composed of 60-element slots. A transition matrix (Figure 3A) was used to sample elements from the ordered-structure to fill each slot. At each slot in the test-sequence the probability of a lag_{+1} transition was constrained to be .7. For example, from Figure 2A, given that element Z was presented, element O was presented next with $p(.7)$. All other lags from the range of 9 lag_{RM} were sampled from the ring with equal probability, $p(.3/9)$. For example, the probability of a specific lag_{RM}, such as lag_{-1}, was $.0\bar{3}$. Figure 3A visually depicts these probabilities. Each lag_{RM} was sampled twice per test-sequence. Twenty-seven 60-element sequences adhering to these transition probabilities were generated for each subject.

To create the interposed generation sequences, the last two test-sequences generated for each subject were modified. Twenty-two sequence elements were pseudo-randomly removed–only elements in list positions six through 60 were eligible for removal and successive elements were not removed–from each sequence and replaced with a “?” cue.

Procedure

For each of the twenty-five SRT trials, Ss were sequentially presented with individual elements from the given 60-item test-sequence in the center of the computer screen. There was a one-to-one element-response mapping and Ss were instructed to respond as quickly and accurately as possible to each element with the appropriate key-press. Once an element was presented it remained on the screen pending a correct key-press. If an incorrect key was pressed the element remained on the screen and a beep sounded. When the correct key was pressed the element was cleared from the screen and the next target appeared immediately. Previous work has suggested 0 RSI is optimal for eliciting implicit learning (Destrebecqz & Cleermans, 2001). Response latencies were measured from element onset to completion of the correct response. If a given RT was $\geq$ 1200ms, an alert screen to "PLEASE RESPOND MORE QUICKLY" was interjected for 500ms into the presentation prior to the onset of the next element. At the completion of each trial there was a break screen until the next trial was initiated by the *S*.

After *S*s had completed twenty-five SRT trials, there were two additional interposed generation trials. *S*s were told that they would again be presented with a stream of sequentially presented elements and were instructed to proceed through the experiment as in the previous twenty-five trials, as quickly and accurately as possible. In contrast to the previous trials, however, *S*s were alerted that they would occasionally see a prompt cue, "?". When presented with a prompt cue, *S*s were instructed to generate the letter that they predicted came next in the sequence. Consistent with the previous twenty-five SRT trials, *S*s were alerted to "PLEASE RESPOND MORE QUICKLY" if RT was $\geq$ 1200ms. Also consistent with the previous twenty-five SRT trials, when presented with an element *S*s were required to respond with the accurate key-press. In contrast to the SRT trials, when presented with a prompt cue, if the *S* pressed any alphabet-character, regardless of the accuracy of the prediction, the *S* was not penalized with a beep and the screen advanced to the next stimulus. If a non-alphabet key was pressed a beep sounded and presentation did not advance until an alphabet-character was pressed. There was a break after the first interposed generation trial, with the second interposed generation trial initiated by the *S*s.

Results and Discussion

Prior to all analyses RT data was subjected to two universal contingencies. First, only accurate responses to letter-elements were included in RT analyses. Second, all subsequent RTs were constrained to be within the interval $[120, 2500]$. The application of these two

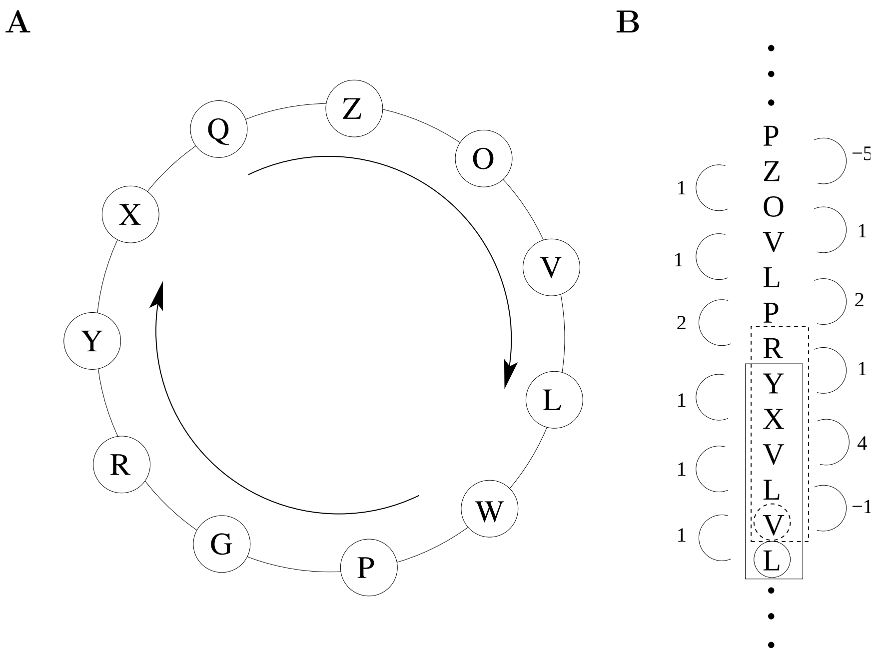

Figure 2: **Elements were arranged into an ordered-structure and test-sequences were generated by probabilistically sampling the structure.** A. Ordered-structure used to generate test-sequences. Elements were randomly assigned a slot in the ordered-structure. Arrows indicate directionality. B. Example test-sequence. Test-sequences were generated by probabilistically sampling elements from the ordered-structure. The numbers indicate lag. Lag is the distance in the ordered-structure between two elements. The sign of each lag indicates direction in the ordered-structure, with clockwise transitions denoted by positive values and counter-clockwise transitions denoted by negative values. The dashed box represents the moving recency-window that captures the elements five time-steps back, relative to element V. The solid box represents the recency-window relative to element L.

universal contingencies reduced the total analyzable data points from 293,345 to 282,676. Beyond the two universal contingencies, there were two separate sections of the experiment, the probabilistic SRT task and the interposed generation task. Each of the two sections of the experiment warranted further data constraints and specific statistical analyses. The results from the probabilistic SRT task are first analyzed, followed by analysis of the results from the interposed generation task.

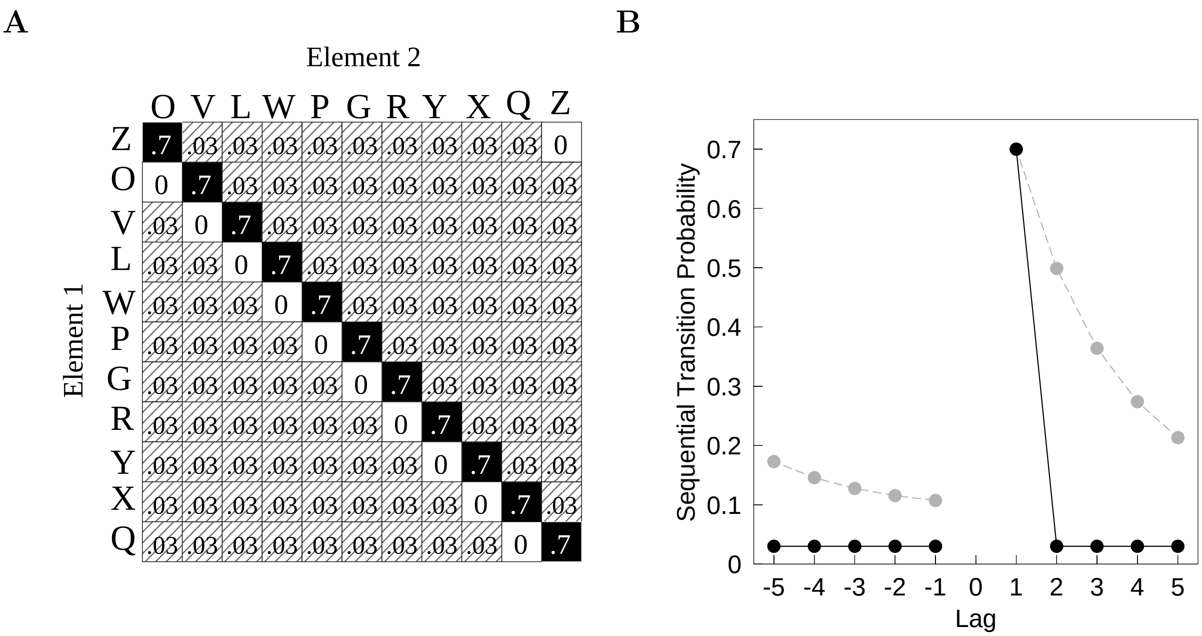

Figure 3: **Test-sequences were created by probabilistically sampling an ordered-structure of elements.** A. Transition matrix. Elements from Figure 2A are decoupled and plotted along the rows and columns of the transition matrix. Each cell of the matrix displays the probability that Element 1 predicted Element 2. The probability of making a lag_{+1} transition was $p(.7)$, and is plotted on the diagonal of the matrix. Each .03 value is a concatenation of (.3/9). B. Lag probability curve. Taken from the transition matrix, the probability of each lag transition at time-step i is plotted in black. Given the high probability of lag_{+1} transitions, there is a non-zero probability of generating a run of elements in a test-sequence composed exclusively of lag_{+1} transitions. For example, from Figure 2A, there is a possibility of generating the sequence Z O V L W P G R Y X Q. Referring back to Figure 3B, the grey points represent the probability that the lag transition at time step i is lag_{+1}, contingent upon all previous lags having also been lag_{+1}, $p(lag_{x_i} = +1 | lag_{x_{i-1}} = +1 ... lag_{x_{i-n}} = +1)$. Lags one through five are equivalent to steps one through five clockwise through the structure. Due to the circular nature of the structure, each lag can be identified by traversing the structure in either the clockwise or counter-clockwise direction. Lags were defined as $[1, 5]$ and $[-5, -1]$, with the element at time-step i defined as lag_0.

Probabilistic SRT Task

Analysis: Learning Curve

The first section of the experiment was composed of trials one through twenty-five and consisted exclusively of the SRT task. Figure 4 plots average RT for the sequence categories, lag_{+1} and lag_{RM}, aggregated across Ss as a function of trial. Examination of

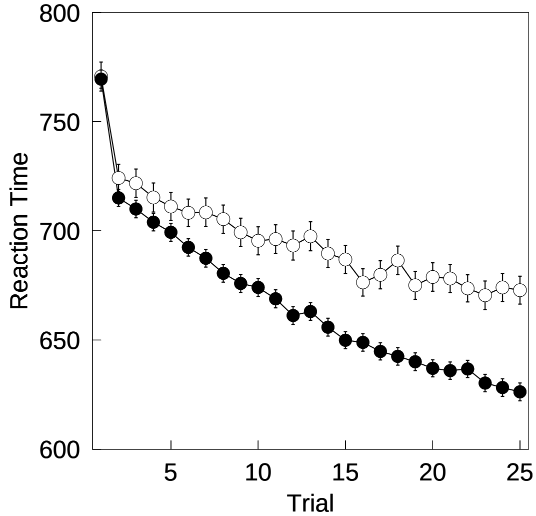

Figure 4: **Learning Curves: Reaction time (RT) is faster to lag_{+1} transitioned elements than to lag_{RM} transitioned elements, and this difference increases as a function of trial.** Accurate-response RTs between 120 to 2500 milliseconds are included in the figure. Error bars reflect 95% confidence intervals. RT was averaged across subjects and plotted for each trial. Closed bullets are RTs to lag_{+1} transitioned elements. Open bullets are RTs to lag_{RM} transitioned elements.

this figure indicates that RT decreases universally across trials. Despite the divergence of RT trajectories with later trials, there appear to be no differences between lag_{+1} and lag_{RM} RTs across the initial trials. Prior to statistical analysis, data were log transformed to minimize deviations from normality that result from the skew of RT distributions. A within-subject repeated-measures ANOVA with log-transformed RT as the dependent measure, trials one through five as regressors and sequence category as a factor (total data points = 56,525), confirmed a highly significant main effect of trial, $F(1, 2042) = 17.1$, Mse = 0.09, $p < .001$, no significant main effect of sequence category, $F(1, 2042) = 0.3$, MSe = .00, $p = .6$, and no significant interaction of trial with sequence category, $F(1, 2042) = 0.2$, Mse = .00, $p = .7$.

The main effect of trial indicates that performance universally improved with exposure to the task. The absence of a main effect of sequence category demonstrates that there were no a priori differences between lag_{+1} and lag_{RM} RTs. Further, the lack of a significant interaction between trial and sequence category indicates that the trajectory of RT increases were not initially influenced by sequence category. Presumably the absence of RT differences across the sequence categories, as well as the lack of an interaction of sequence category and trial, reflect initial skill-learning of the SRT task and corresponding element-response mappings.

Consider again Figure 4. After initial task and stimulus familiarization, Figure 4 illustrates that although RT continued to decrease universally as a function of trial, RT was differentially affected by sequence category. Relative to lag_{RM} transitioned elements, RTs were faster to lag_{+1} transitioned elements. Moreover, the advantage for lag_{+1} transitioned elements appears to increase as a function of trial, widening the performance gap between the two sequence-category RT trajectories. A within-subject repeated-measures ANOVA with log-transformed RT as the dependent measure, trials six through twenty-five as a regressor and sequence category as a factor (total data points = 262,151) demonstrated a significant main effect of trial, $F(1, 8192) = 41.9$, MSe = .2, $p < .001$, a significant main effect of sequence category $F(1, 8192) = 48.7$, Mse = .3, $p < .001$, and a significant interaction of trial with sequence category, $F(1, 8192) = 4.4$, Mse = .02, $p < .05$.

The main effect of trial indicates that RT performance continued to improve for both sequence categories with continued exposure. The main effect of sequence category and the interaction of sequence category and trial demonstrate that beyond mere element exposure, RTs to probabilistic lag_{+1} transitioned elements were differentially facilitated and this advantage grew with learning. In short, the main effect of sequence category and the interaction of sequence category and trial suggests that subjects became increasingly sensitive to the probabilistic statistics of the test-sequence.

Analysis: CRL Curves

The conditional response latency statistic was used to quantify learning across this section. The conditional response latency is a RT measure that parses RT as a function of lag

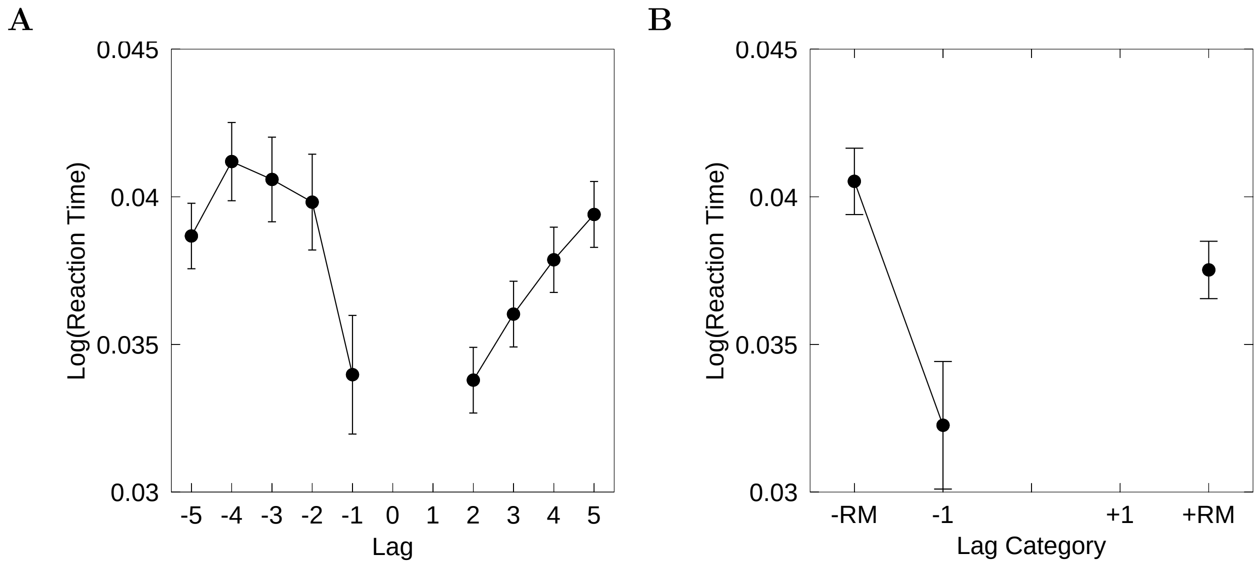

Figure 5: **CRL Figures: while probability of** lag_{RM} **transitions were equivalent across lag,** lag_{-1} **RTs were faster than** lag_{-RM} **RTs, and** lag_{+RM} **RTs increased as a function of lag.** Trials one through five were excluded from analysis, as well as the initial five RTs of each trial. Only accurate-response RTs between 120 to 2500 milliseconds are included in the figures. RTs were log-transformed and then normalized by the averaged RT of lag_{+1} transitions by subject and trial. Error bars in both panels reflect 95% confidence intervals. A. Conditional response latency across all lags. Points represent average RT plotted as a function of lag. Lag_{+1} was equivalent to zero as a function of the normalization process so is not plotted. B. Conditional response latency across lag categories. Lag_{RM} was partitioned into three categories: lag_{-1}, lag_{+RM} and lag_{-RM}. The lag_{+RM} category included lags $[2, 5]$ and the lag_{-RM} category included lags $[-2, -5]$.

(lag-CRL). Lag, in this context (for inter-response time applications see Kahana & Loftus, 1999; Howard & Kahana, 1999; Murdock & Okada, 1970; Zaromb et al., 2005; Kahana & Howard, 2005; Kahana et al., 2008), is defined as the distance in the ordered-structure between a stimulus presented at time-step i and the stimulus presented at time-step $i + 1$.

Prior to lag-CRL analysis, the data was subjected to three additional constraints. First, trials one through five were regarded as preliminary and were excluded from subsequent lag-CRL analysis. Additionally, the first five responses from each trial were removed. Second, a pilot analysis demonstrated that RTs to accurate responses immediately following inaccurate responses were artifactually facilitated. For example, if when presented with element A, the S incorrectly pressed F, a beep sounded and the S

was required to press the correct key to advance forward in the sequence. The RT to the correct key, A in this example, was facilitated from having been primed and was therefore excluded from subsequent lag-CRL analyses. The third constraint to the data was the instantiation of a five-item recency-window: an element was excluded if it had appeared within five elements of its previous presentation. For example, lag$_{-1}$ transitions overwhelmingly occurred as part of sequences following the presentation format A B A. It is therefore possible that robust backward associations could result from a simple priming effect. To control for priming effects, the recency-window excluded elements if they had appeared within the previous five element-responses.

In summary: data was constrained to include only accurate responses, all RTs were constrained to be within the interval $[120, 2500]$, the first five trials of the experiment were considered practice and removed, the first five 'warm-up' responses of each trial were further excluded from analysis, responses immediately following inaccurate responses were excluded, and a five-element recency-window was instated. The resultant data was then log transformed to minimize deviations from normality that result from the skew of RT distributions. The application of these constraints resulted in 158,739 analyzable data points.

To account for individual differences inherent in between-subject designs, the lag$_{RM}$ sequence category data was normalized. Average lag$_{+1}$ $log(RTs)$ were calculated as a function of S for each trial and subtracted from subject- and trial-matched lag$_{RM}$ RTs. Normalized RT values greater than zero represent RTs slower than the average lag$_{+1}$ RT, zero values indicate performance equivalent to the average lag$_{+1}$ RT, and values less than zero indicate RTs faster than the average lag$_{+1}$ RT.

Figure 5A plots the lag-CRL for lag$_{RM}$ across Ss and trials. Importantly, lag$_{RM}$ sequenced elements do not reflect the underlying probabilistic structure of the lag presentation. That is, despite the fact that lag$_{RM}$ transitioned elements were all equally likely to be presented, RTs appear to vary systematically, with RT increasing as a function of lag, particularly for the lag_{+RM} category. While Figure 5A represents data collapsed across trials, data at the individual S by trial level was rather unbalanced across the various response categories: there were missing response-values for individual lags on any given trial across Ss due to the conservative inclusion restrictions imposed upon the data set. Therefore, to ensure that all Ss were represented across all lags, trials were grouped into 4 blocks: trials 6-10; trials 11-15; trials 16-20; and trials 21-25. An initial within-subject repeated-measures ANOVA with normalized log-transformed lag_{RM} RT as the dependent measure, lag as a factor and block as a regressor, confirmed a highly significant main effect of lag, $F(8, 39774) = 4.18$, Mse $= 0.03$, $p < .001$, a highly significant main effect of trial series, $F(1, 39774) = 398.54$, MSe $= 2.59$, $p < .001$, and no significant interaction of lag with trial series, $F(8, 39774) = 1.42$, Mse $= .01$, $p = .18$.

The main effect of lag indicates that some subset of lags from within the lag_{RM} category produced differentially faster RTs relative to the other lags, as visually depicted in Figure 5A. The main effect of trial series indicates that RT performance continued to improve for all lags with continued exposure. The lack of an interaction of lag and trial series demonstrates that the average RT differences across the lag_{RM} category did not differ significantly across the blocks of trials.

From the main effect of lag, rather than a factorial set of comparisons, the specific interest was in determining if there was a reliable lag_{-1} effect, as well as determining if there were statistically reliable graded RTs as a function of increasing lag. Evidence for a lag_{-1} effect would include significant RT differences between lag_{-1} transitioned elements and other lag_{RM} transitioned elements. Evidence for lag-mediated graded RTs would include statistically significant RT differences between adjacent lags, such as the comparison lag_2 to lag_3. Therefore, separate analysis were specifically conducted on each of the sets of lags: lag_{-1}, lag_{-RM}, lag_{+RM}, with the first analysis the comparison of the averages of the three groups of lags.

Analysis: lag groups. The omnibus ANOVA with individual lags as factors confirmed the main effect of lag suggested by Figure 5A . To examine this effect, the lag_{RM} category was divided into three groups: lag_{-1} (total data points = 1,670); lag_{-RM}, which consisted of lags $[-2:-5]$ (total data points = 16,053); and lag_{+RM}, which consisted of lags $[2:5]$ (total data points = 22,273). Figure 5B plots the normalized log(RTs), averaged by S across lag category and trial series. The figure demonstrates two important points. First, RT to lag_{-1} transitioned elements was faster than RT to either category of lag_{RM} transitioned elements. Second, lag_{+RM} RTs appear to be significantly faster than the lag_{-RM} RTs.

A within-subject repeated-measures ANOVA with normalized log-transformed RT as the dependent measure, lag group as a factor and trial series as a regressor, confirmed a main effect of lag group, $F(2, 408) = 10.03$, Mse = 0.01, $p < .001$, a main effect of trial series, $F(1, 408) = 112.5$, MSe = .19, $p < .001$, and no significant interaction of group with trial series, $F(2, 408) = 0.97$, Mse = .00, $p < .38$. The main effect of group demonstrates significant differences between at least two of the groups, as suggested from Figure 5B. The main effect of trial series indicates that RTs decreased as a function of trial series. The insignificant interaction of lag group and trial series indicates that the difference between lag groups was not significantly different across trials.

To specifically examine the main effect of group, Wilcoxon rank-sum tests with Bonferroni's correction were performed between lag categories across trial series for lag_{-1} and lag_{+RM}, lag_{-1} and lag_{-RM}, and lag_{+RM} and lag_{-RM}. As suggested by Figure 5B as well as the main effect in the corresponding ANOVA, lag_{-1} RTs were significantly faster than both lag_{+RM} ($p < .001$) and lag_{-RM} ($p < .001$) RTs. Additionally, lag_{+RM} RTs were significantly faster than lag_{-RM} RTs ($p < .001$).

Analysis: lag_{-RM}. Inspection of Figure 5A indicates that there is no visual evidence for graded RTs as a function of lag across the lag_{-RM} category. A within-subject repeated-measures ANOVA with normalized log-transformed RT as the dependent measure, lag_{-RM} as a factor (i.e, lags $[-2, -5]$) and trial series as a regressor confirmed no main effect of lag, $F(3, 15841) = 0.68$, Mse $= 0.00$, $p < .55$, a main effect of trial series, $F(1, 15841) = 156.0$, MSe $= .03$, $p < .001$, and no significant interaction of lag with trial series, $F(3, 15841) = 1.26$, Mse $= .01$, $p = .29$. In short, the lag_{-RM} RTs did not differ significantly across lags. While lag_{-RM} RTs did decrease with trial, the increases were not significantly different across the category.

Analysis: lag_{+RM}**.** The statistical environment of stimulus presentation was designed such that all lag_{RM} transitions, both forward and backward, were constrained to the predictive probability $p(.3/9)$. Therefore, graded contiguity effects, which would be evidenced by RT increases as a function of lag, would not be a reflection of the underlying statistics of the probabilistic environment.

Averaged across S by trial, the lag_{+RM} RTs plotted in Figure 5A appear to increase gradually across lags. To examine lag_{+RM}, data was first submitted to a within-subject repeated-measures ANOVA with normalized log-transformed RT as the dependent measure, lag as a factor and trial series a regressor. Results confirmed a main effect of lag, $F(3, 22061) = 5.19$, Mse $= 0.03$, $p < .01$, a significant main effect of trial series, $F(1, 22061) = 217.13$, MSe $= 1.4$, $p > .001$, and no significant interaction of group with trial series, $F(1, 22061) = 1.89$, Mse $= .01$, $p = .13$.

The main effect of lag from the omnibus lag_{+RM} ANOVA indicates significant differences among at least two of the lags included in the lag_{+RM} category. The main effect of trial shows that as Ss became sensitive to the probabilistic sampling environment, RTs to lag_{+RM} elements decreased across trials. The lack of an interaction between lag and trial series indicates that the average RT differences across the lag_{+RM} category did not differ significantly across the blocks of trials.

To test for graded associations across the lag_{+RM} category, a simple linear regression

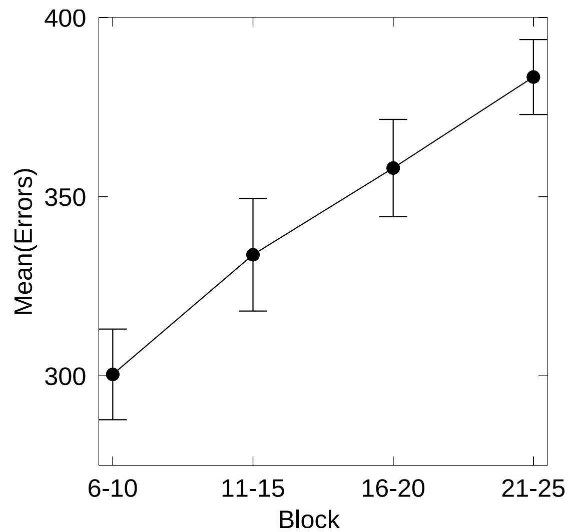

Figure 6: **Error Rate: average total number of errors increased as a function of block.** Trials one through five were excluded from analysis, as well as the initial five responses of each trial. Only inaccurate-responses with RTs between 120 to 2500 milliseconds are included in the figure. Five-item recency window was applied to the data. Error bars reflect 95% confidence intervals. The number of errors was summed across subjects for each trial; trials were then collapsed into blocks consisting of trials 6-10, 11-15, 16-20 and 21-25.

was performed with normalized log-transformed RT as the dependent measure and lag as a regressor. Results confirmed evidence for a linear relationship across RT as a function of lag, $F(1, 22271) = 14.22$, $p < .001$. There was a significant positive linear relationship between RT and lag: $rate = 0.03 + (.001)lag, R^2 = .0006$. As forward-going lag-values increase, RT increases, a result that supports the hypothesis that RTs should increase as a function of increasing lag. However, only .06% of the variation in RT is due to differences in lag_{+RM}, such that lag_{+RM} is not a good predictor of RT. This result is consistent with the low prediction statistic ($p(.3/9)$) associated with each member of the lag_{+RM} category.

Interposed Generation Task

Error Rates in SRT Task

Recall that all of the analyses of the SRT data were constrained to include only accurate responses. However, in so far as Ss became increasingly sensitive to the underlying ordered-structure from which elements were probabilistically sampled, error rates could potentially vary systematically across blocks. Figure 6 displays the mean error rates collapsed across S for each block of trials. From visual inspection of the figure, error rates do appear to increase with exposure to the probabilistic environment. That is, with practice, while RT decreased universally (Figure 4), as well as systematically as a function of lag (Figure 5), the average error rate actually increased. This pattern of results could be consistent with a general sensitivity to the probablistic learning environment, or, it could also be a reflection of explicit awareness of the underlying sequence. To specifically examine if Ss were explicity aware of the underlying sequence, Ss were tested on a generation task.

Analysis

The second section of the experiment was composed of trials twenty-six and twenty-seven. Procedurally identical to the SRT task in trials one through twenty-five, Ss were required to respond to elements presented on the screen with the appropriate keypress. However, in addition to elements from the ordered-structure, generate-response cues were interposed in place of some of the elements. When presented with the prompt cue, "?", Ss were required to generate a prediction by pressing the key representing the element that they predicted should appear next. The dependent measure in the interposed generation task was the frequency of the S-generated lags.

Proportions were used to measure learning across this section. A proportion is a frequency statistic that quantifies the magnitude of selected subsets of data in comparative relation to the whole data set. Proportions enabled the comparison of the data subsets that were a priori of empirical interest in the interposed generation task. To calculate a proportion statistic for the interposed generation data, a group of lags was first selected from the total data set. For example, consider the lag_{+RM} group. The lag_{+RM} group is composed of the lags lag_{+2}, lag_{+3}, lag_{+4}, and lag_{+5}. To calculate the proportion of each lag relative to the lag_{+RM} set, the frequency of each lag becomes a numerator, and the sum of the numerators becomes the denominator. Each numerator statistic is therefore a proportion of the denominator, such that the numerators form a conditionalized probability distribution that sums to one.

Recall that all data was subjected to two universal contingencies. First, only accurate responses to letter-elements were included in analyses. If *S*s made an incorrect response to a letter-element, for example if the *S* was presented with the element "R" and the *S* pressed the key "L", this response was removed from analysis. In contrast, when presented with the generate cue-element, "?", any alphabet response a *S* generated was included in the analysis. The second universal data contingency was that RTs were constrained to be within the interval $[120, 2500]$. The RT interval was imposed for all data, that is, both the letter-element and the cue-element responses. While the dependent measure in the interposed generation task was the frequency of *S*-generated lags, the RT range was applied to ensure that *S*s were adhering to the instructions of the SRT task.

Prior to analysis, the interposed generation data was subjected to two further constraints. First, consistent with the SRT data from trials six through twenty-five, the first five responses from each trial were removed. Second, again consistent with the SRT data from trials six through twenty-five, a five-item recency-window was applied. The recency-window excluded from analysis any element that had been presented or generated within five elements of its previous presentation. For example, consider the sequence in which the *S* is prompted to generate an element at serial position five: A B C D ? G A B C. Now assume that the *S* generated the element C in response to the generate cue, A B C D **C** G A B C. The recency window would exclude the generated C from analysis because this element had been presented within five items, at serial position three within the sequence. Additionally, element C in serial position nine of the sequence would also be excluded from analysis, due to the generated C at serial position five.

In contrast to the constraints applied to the SRT data in section one of the experiment, there were three constraints that were relaxed for the interposed generation data analysis. First, as previously noted, while response data in the condition in which letter-elements were presented (e.g., A,D,K...) continued to be constrained to include only correct responses, response data in the condition in which the cue-element was presented (i.e., "?") necessarily included both sequential, lag_{+1} responses, as well as non-sequential, lag_{RM} responses. Second, data was not normalized by lag_{+1} RTs; the interest in the interposed generation task was the frequency of S-generated lags, not differences in RT to letter- vs. cue-element responses. Third, accurate responses to letter-elements immediately following inaccurate responses to letter-elements were not excluded from analysis. While preliminary analysis in the SRT task in section one of the experiment indicated that accurate responses immediately following inaccurate responses were artifactually facilitated, the goal of the interposed generation task was to compare proportions of S-generated responses, not RTs. The application of data constraints reduced the total data points from 27,968 to 25,798.

To increase the pool of predictions for each S, the interposed generation trials were collapsed across trials, which collapsed the available 25,798 data points to 2,461. S-generated letters that were not part of a S's 11-element pool were labeled extra-list intrusions (eli). Predictions were grouped into three categories: lag_{+1}, lag_{RM} and eli. A primary comparison of the proportions of the three categories was first performed. Specifically, a proportion for each of the three prediction categories was calculated for each S by dividing the total number of predictions in each category by the total number of predictions summed across all three prediction categories. Figure 7A plots the proportion of S-generated predictions for the lag_{+1}, lag_{RM} and eli categories. Plotted beside the proportion of generated predictions is the probability of lag_{+1} and lag_{RM} transitions ($p(eli) = 0$) from the probabilistic SRT task of trials one through twenty-five. From the figure, in contrast to the prediction statistics of the SRT task, Ss appear to have generated significantly more lag_{RM} predictions than lag_{+1} predictions. A paired Wilcoxon rank-sum test was performed between the lag_{+1} and lag_{RM} categories and results confirmed that Ss did generate significantly more lag_{RM} than lag_{+1} predictions ($p < .001$).

To examine the lag_{RM} category, proportions were generated for each individual lag_{RM} lag, relative only to the lag_{RM} category. Figure 7B plots the proportion of generated responses, averaged across Ss, as a function of lag. From the figure, lag_{-1} elements appear to be predicted more than other lag_{-RM} elements, and lag_{+2} elements appear to be predicted more often than all other lag_{RM} elements, including lag_{-1} elements. A within-subject repeated-measures ANOVA with conditionalized lag_{RM} probability as the dependent measure and lag as a factor, confirmed a main effect of lag, $F(8, 1616) = 7.93$, Mse $= 0.10$, $p < .001$. The main effect of lag demonstrates significant differences between at least two of the lags, as suggested from Figure 7B.

While there is a main effect of lag in the lag_{RM} analysis, given the smaller data set generated from the two-trial interposed generation task, the five-item recency-window differentially affects the lag_{-RM} and lag_{+RM} responses. That is, the recency-window removes a full 56% of responses from the lag_{-RM} group, in comparison to a lesser 42% of responses from the lag_{+RM} group. Therefore, analysis was restricted to comparisons within the lag_{-RM} and lag_{+RM} groups, respectively.

To specifically examine the lag_{RM} groups, a Wilcoxon rank-sum test with Bonferroni's correction was performed between the lag_{-1} and lag_{-RM} groups, and the lag_{+2} and lag_{+RM}-modified groups. As suggested by Figure 7B, the proportion of lag_{-1} generated responses was significantly greater than the proportion of lag_{-RM} ($p < .002$) generated responses. Again from Figure 7B, the proportion of lag_{+2} S-generated responses appears to be greater than the proportions of the positive lag responses. For these analyses, lag_{+RM}-modified was defined as lags ranging from $[3:5]$. A Wilcoxon rank-sum test with Bonferroni's correction confirmed that the proportion of lag_{+2} S-generated responses was significantly greater than the proportion of lag_{+RM}-modified S-generated responses ($p < .001$).

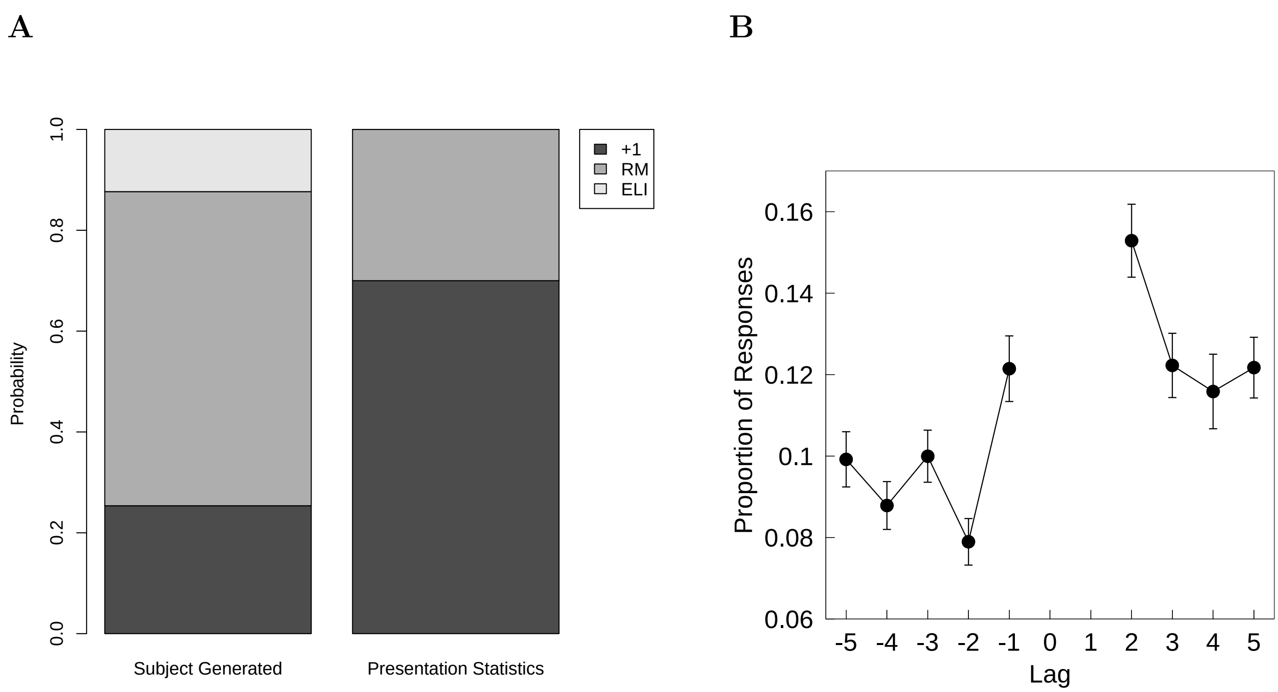

Figure 7: **Interposed Generation Task: Ss generated more** lag$_{RM}$ **elements than** lag$_{+1}$ **elements. Given that Ss generated a** lag$_{RM}$ **element, Ss generated more** lag$_{-1}$ **elements than** lag$_{-RM}$ **elements, and more** lag$_{+2}$ **elements than any other** lag$_{+RM}$ **element.** Excluded from analysis were the initial five stimuli on both trials, stimuli in which response RTs either fell below 120ms or exceeded 2500ms, and stimuli that had been presented within a five-item recency-window. Trial was collapsed across Ss. A. Proportion of lag$_{+1}$, lag$_{RM}$ and eli (extra-list intrusion) generated elements plotted beside the proportion of lag$_{+1}$ and lag$_{RM}$ transitions ($p(eli) = 0$) from the probabilistic SRT task of trials one through twenty-five. B. Proportion of each lag$_{RM}$ generated element as a function of the total number of lag$_{RM}$ elements generated. Proportions were calculated by S for each lag by dividing individual lags by the total number of lag$_{RM}$ elements generated. *Lag* quantifies the distance from the preceding to the measured stimulus in the latent stimulus-structure. The sum of the proportions across lags is one. Error bars reflect 95% confidence intervals.

General Discussion

The goal of the experiment was to examine and quantify sensitivity to backward and graded associations in a non-episodic memory task. While backward and graded associations are a ubiquitous finding across multiple episodic memory paradigms, these associations had not been specifically examined in the context of the statistical learning paradigm. Results from the probabilistic SRT task indicated that Ss did make backward associations, as well as forward-graded associations. To assess explicit sequence knowledge, after the probabilistic SRT task Ss completed two additional trials in which they were required to generate a prediction at various points throughout the sequence. The distribution of generated responses from the "interposed generation task" did not mirror the presentation statistics of the probabilistic SRT task, which suggests that Ss were not able to explicitly produce the underlying sequence order. The shape of the associations demonstrated in the experimental data, with emphasis on the similarity to the shape of associations evidenced in episodic memory tasks, is atheoretically discussed below.

RT Differences and Priming Effects

The primary dependent measure in this test of sequential learning was RT differences among the probabilistically equivalent lag_{RM} transitions. However, the interpretation of RTs across Ss can be complicated by individual performance differences. That is, lag_{-1} and lag_{+3} RTs cannot be unambiguously compared between a faster S and a slower S due to individual differences across baseline RTs (Chapman, Chapman, Curran, & Miller, 1994). To control for individual RT differences, each S's data was log-transformed and then normalized by the S's average lag_{+1} RT, as a function of trial.

$$\text{RT}_{\text{normalized}} = \log(\mu_{\text{lag}_{+1}}) - \log(\text{RT})$$

In addition to individual S differences, priming effects can also drive artifactual RT differences between lags. Importantly, the lag_{-1} category was particularly subject to priming effects. As an example consider the sequence A B A. RT could be faster to the second presentation of A, relative to the first presentation of A, by virtue of the fact that A had just recently been presented. Any conclusions that a backward effect was present in the A B A data would necessarily be distorted by the presence of a simple priming effect. To control for priming effects in the experiment, a recency-window was instated in the analysis: elements were excluded from analysis if they had appeared within the previous five element-responses. The recency window necessarily eliminated a great number of transitions, particularly from the lag_{-RM} and lag_{-1} categories. While the design of the experiment allowed elements that deviated from the underlying sequence-order to be presented within a five-item recency window, the design of the analysis was conservative enough to enable detection of only robust lag effects.

Backward and Graded Associations

Backward associations are a ubiquitous finding within episodic memory (for a review, see Kahana et al., 2008). Episodic memory recall tasks include free recall, serial recall and paired-associate learning. In the context of episodic free and serial recall tasks, both backward and graded associations are beneficial to the extent that they increase the probability of recalling sequence elements. In contrast, backward and graded associations are anti-predictive in paired-associate learning tasks. None the less, backward and graded associations are formed during paired-associate learning, with backward associations the primary source of interference (Primoff, 1938; Umemoto & Hilgard, 1961; Young, 1961; Slamecka, 1976; Howard et al., 2009; Provyn et al., 2007). The phenomenon that backward and graded association have been demonstrated in data across episodic memory tasks, despite the fact that the associations may inhibit performance in specific tasks, leads to the question of whether backward and graded associations are a natural consequence of extracting regularity in structured environments, regardless of the surface-level task.

Both episodic memory tasks, such as free and serial recall, and statistical learning tasks, such as the probabilistic SRT task, require learners to extract regularity from a structured environment. If similar associative properties are present in both episodic and statistical learning, and backward and graded associations are formed during episodic learning, then backward and graded associations may also be formed during statistical learning. The lag_{RM} category, composed of probabilistically equivalent lags, was specifically designed to detect the formation of associations in the context of a statistical learning task, that would not be beneficial for predicting the most statistically predictable lag_{+1} transition.

From the probabilistic SRT task data, Ss did demonstrate significantly faster RTs to lag_{-1} elements than to other lag_{-RM} elements. In other words, Ss displayed differential sensitivity to the lag_{-1} transition, despite the fact that this transition was no more predicted than any other lag_{-RM} element. Moreover, the lag_{-1} effect is arguably robust given that the five-item recency-window significantly reduced the number of instances of lag_{-1} transitions (n=1670), relative to other lag_{-RM} transitions (n=16053).

Conditional Probability

Joint and conditional probabilities are often employed to describe the relationship between elements in the context of statistical learning. Joint probability signals the overall frequency with which two elements co-occur, $p(\text{X}, \text{Y})$. Conditional probability measures the frequency of one element given another element, $p(\text{Y}|\text{X})$, and has predictive power in tests of sequential learning. Importantly, while X and Y can have a high joint probability, they will have a low conditional probability if one of the elements does not routinely predict the other. The role of joint and conditional probability has been examined in both auditory (Saffran et al., 1997) and visual (Fiser & Aslin, 2001) statistical learning, with results indicating that statistical learners are sensitive to conditional probability statistics, even when joint probability statistics have been equated.

While elements in the probabilistic SRT task of the experiment were presented sequentially such that there were no true joint probability co-occurrence statistics, each lag_{RM} did have a specifically controlled conditional probability: each lag_{RM} transition was sampled with $p(.3/7)$. However, as previously noted, the conservative recency-window differentially reduced the instances of particular lags. Consider an example of an illegal and a legal sequence containing a lag_{-1} transition. The sequence A B A contains the lag_{-1} transition from B to A, which would be excluded by the recency-window. The sequence G B A contains the lag_{-1} transition from B to A, which would not be excluded by the recency-window. In short, while lag_{-1} transitions were defined in the experimental design to be sampled with $p(.3/7)$, the analysis constrained the conditional probability to be a subset of the sampled transitions.

As illustrated in the example, legal lag_{-1} transitions were marked by a "jump" in the underlying ordered-structure: the sequence "jumped" from G to B (lag_{-5}) then back from B to A (lag_{-1}). These jumps in the sequence order should have presumably yielded slower RTs to lag_{-1} transitioned elements. None the less, *S*s demonstrated significantly faster RTs to lag_{-1} transitioned elements than to lag_{-RM} transitioned elements. The lag_{-1} association is therefore arguably robust and fairly remarkable given the subset of data included in the analysis. Similar to the results demonstrated in the paired-associate learning paradigm, backward and graded associations appear to have been formed automatically, regardless of whether this property was beneficial or relevant to the surface-task of prediction.

Chaining

While each lag_{RM} transition was constrained to be equivalently predictive, the high probability of lag_{+1} transitions gave rise to runs of a series of lag_{+1} transitions within the test-sequences. Figure 3B visually displays the probability of generating a run of elements in a test-sequence composed of exclusively lag_{+1} transitions at each slot in the ordered-structure. From the figure, despite the equivalent sampling probability of each lag_{RM} transition (solid black lines) there is an appreciable degree of predictability of an element $n + i$ at time-step i (solid grey lines) due to the presence of longer-range predictive relationships. These longer-range predictive relationships are particularly marked in the forward direction and necessarily decrease a function of lag. In episodic recall tests of serial learning, in which each element of a series elicits the next element, elements can be associated together in chains of larger chunks of elements (Tichner, 1909). The result that Ss demonstrated forward-graded associations could be a reflection of Ss' sensitivity to these longer range “chains” of conditional probabilities.

While the forward-graded associations might be mediated by the longer-range forward-conditional probabilities of traversing the underlying-ordered structure, the lag_{-1} association is less easily accounted for. The probability of traversing the entire structure to time-step $i = 10$, contingent upon all previous lags having been lag_{+1}, $p(lag_{x_{i=10}} = +1|lag_{x_{i-1}} = +1...lag_{x_{i-9}} = +1)$, is necessarily lower than any other run of lag_{+1} transitions through the sequence. Given the fact that all remote lags, of which lag_{-1} is a member, are equally likely to have occurred with probability $p = .03$, and given the fact that the probability of transitioning sequentially through the entire 11-element ordered-structure is quite low $(p(.107))$, the lag_{-1} transition yields an element that is statistically the least predictive in this probabilistic environment. Nonetheless, Ss responded significantly faster to lag_{-1} elements than to other lag_{-RM} elements.

Temporal Contiguity

Each element and its backward associate were consistently presented close together in time by virtue of the high probability of lag$_{+1}$ transitions. For example, from the sequence A B C D element B and its backward associate, element A, were presented in close temporal proximity on approximately 70% of presentations. Temporal contiguity effects refer to the phenomenon that associations are formed between elements presented close together in time. For example, in tests of free recall in which *S*s recall a list of words in the order they come to mind, the probability of making a recall transition from a just-recalled word to other words in the list is higher for words originating from nearby serial positions, relative to remote serial positions (for review, see Kahana et al., 2008). In contrast to episodic memory tasks such as free recall, simply associating elements together by virtue of close temporal proximity will not yield accurate prediction statistics in the statistical learning SRT task. The direction of the association must also be encoded for prediction. The presence of the lag$_{-1}$ effect, which is actually under-sampled and the least predictive transition in this prediction task, is therefore rather paradoxical. Associations beyond those that are statistically most probable are anti-predictive in an SRT task. The experiment provides evidence that properties of associative learning are not necessarily task-mediated. The lag$_{-1}$ association, and to a lesser extent the forward-graded associations, may reflect a general binding of the elements occurring in close temporal contiguity, irrespective of the predictive relationship between the elements. Given that statistical learning is a kind of associative learning, and given that temporal contiguity is a mechanism of associative learning that has been used to describe the same associative properties demonstrated across episodic memory tasks, the lag$_{-1}$ and forward-graded associations in this experiment are consistent with general properties of associative learning.

Chunking and Hierarchic Coding

In addition to conditional transition probabilities and temporal contiguity, statistical learners may attend to multiple sources of information simultaneously in an effort to reduce environmental uncertainty and detect regularity in structured input streams. Consider the following three examples of the interplay between bottom-up segmentation (i.e., transitional probabilities) and top-down lexical segmentation, which support the notion that the statistical learner can combine statistical cues from the structured input with other segmentation cues.

First, previous experience with language stress patterns has been demonstrated to shape infant statistical learning. When presented with streams of unfamiliar words, infants demonstrate facilitated word-segmentation performance if the unfamiliar words match the infants' native language stress patterns (Thiessen & Saffran, 2003). Second, in addition to stress-patterns, pre-exposure to either disyllabic or trisyllabic words can induce a word-length expectation in infants. That is, infants pre-exposed to nonsense words inconsistent with the length of words embedded in fluent speech during a segmentation task were unable to discriminate words from part-words (Lew-Williams & Saffran, 2012). And third, recently acquired "anchor" words can facilitate adult *S*s' ability to segment words in a new language when the recently acquired words appear in the continuous stream of input (Cunillera, Camara, Laine, & Rodriguez-Fornells, 2010). All three of these examples provide evidence that statistical learning is not an isolated mechanism. Instead, previous experience can induce a prior, or learning bias, that shapes the ability to process subsequent sequential input.

Certainly conscious rehearsal processes are associative mechanisms that are utilized in tests of episodic learning, in which lag$_{-1}$ and graded-associations have been widely demonstrated. It is possible that the lag$_{-1}$ and forward-graded associations present in this data reflect a combination of sensitivity to chains of conditional probabilities, temporal contiguity, and conscious rehearsal processes. For example, sequences of FOC could be learned by memorizing chunks of bigrams and trigrams, as proposed by proponents of exemplar-based models (Perruchet & Pacteau, 1990; Servan-Schreiber & Anderson, 1990). Chunking in this context would involve explicitly remembering high-frequency fragments within the sequence, such as runs of lag_{+1} transitions. It is possible that higher-order conditionals build upon lower-level conditionals. That is, FOCs could be embedded in SOCs, such that learners chunk adjacent dependencies and then form higher-order relations between chunks of FOCs, and these relations result in forward-graded associations.

Consider the Nissen and Bullemer (1987) sequence 4-2-3-1-3-2-4-3-2-1. A chunking strategy for learning the sequence would be to segment the longer sequence into smaller chunks, such as 4-2-3, 1-3-2, ... or 4-2, 3-1, 1-3... . Runs of lag_{+1} transitions from the ordered-structure (e.g., Figure 3B) provide an environment that could be supported by a chunking strategy. A hierarchic representation strategy simply extends the chunking mechanism to represent chunks at multiple levels, including: the entire sequence (e.g., 4-2-3-1-3-2-4-3-2-1); chunks of the sequence (e.g., 4-2-3, 1-3-2, ... or 4-2, 3-1, 1-3...); and individual elements within the chunks (e.g., 4, 2, 3 ...).

To measure whether Ss were explicitly aware of the underlying stimulus-structure in this experiment, cues requiring a S-generated prediction were interposed into 2 additional trials of the SRT task. This "interposed generation task" was used to evaluate the probability of specific learner-generated lag-transitions to the prompt cues. If Ss were explicitly aware of the underlying sequence order, then when prompted to generate the next response in the sequence, the proportion of lag_{+1}-predicted elements should approximate the probabilistic sampling statistic of the SRT trials ($p(.7)$). Results indicated that Ss did not generate the presentation statistics of the SRT task: Ss generated a greater proportion of lag_{RM} elements than lag_{+1} elements. In fact, given that Ss generated a lag_{-RM}, the proportion of lag_{-1} transitions was greater than other lag_{-RM} transitions. This result is similar to the finding from the paired-associate learning literature: given that an incorrect element is recalled, the lag_{-1} transition is generated significantly more than other lag_{RM} transitions.

The differing pattern of results from the probabilistic SRT data and the interposed generation data may indicate that while Ss were sensitive to the underlying ordered-structure from which the sequences were composed, this learning was not demonstrable in an explicit variant of the SRT task. The interposed generation task cannot be used to conclusively reject the notion that Ss had some explicit knowledge of the underlying ordered-structure. What the task does demonstrate is that Ss were not able to explicitly generate the the probabilistic presentation statistics of the SRT task. Despite the inability to explicitly generate this associative learning, the properties of associations across the lag_{-1} and lag_{+RM} elements in this SRT task is consistent with the properties of associations across episodic recall tasks.

Associative Learning

Statistical learning is simply the name given to the domain general process of extracting variability in the environment to reduce uncertainty and, correspondingly, make statistically beneficial predictions. The SRT task provides a device for controlling, quantifying and examining transitional probabilities among adjacent and non-adjacent elements in structured sequences. If statistical learning is to be considered a paradigm that is unrelated in terms of learning mechanisms to episodic learning, then the examination of the associative properties of the SRT data from this experiment yielded paradoxical associations. The fact that similar associative properties demonstrated across episodic learning are also present in this data could be evidence for the presence of episodic learning that then contaminated the data. However, results from the interposed generation task indicate quite the contrary: there was not substantive evidence for episodic learning of the underlying statistical structure of the data.

There is no a priori reason that associations beyond the most statistically predictive associations are beneficial in statistical learning tasks. In fact, associations beyond the most statistically predictive are arguably anti-predictive. If the task is to anticipate the next most likely element given the conditional probabilities of the distribution of elements, and if a lag_{-1} transition is not statistically likely, then this transition should not be anticipated. This paradoxical prediction of the past in the context of this experiment is not, however, inconsistent with work on language learning. The ability to extract sequential structure in the environment is crucial for word segmentation, and sensitivity to backward transitions has been demonstrated to be beneficial toward this end (Aslin et al., 1998; Willits et al., 2009; Pelucchi et al., 2009).

Backward associations appear to be a ubiquitous property of associative memory that transcend surface level task characterstics. These associations are present in and beneficial for word segementation and recall tasks. These associations are not however limited to recall tasks nor to tasks in which they are necessarily beneficial. Backward associations are present in cued recall data, despite being the largest source of interference, and in the context of the SRT task are actually anti-predictive. What appears to be a common denominator for the creation of lag_{-1} associations is not necessarily the task itself, nor the paradigm under which a task is defined. Rather, what appears to be necessary for the creation of lag_{-1} associations is statistically structured input. If statistical learning is to be considered a paradigm with shared or at least similar learning mechanisms as associative learning paradigms, then the associative properties emergent from the structured-SRT data from this experiment, while anti-predictive, were in fact predictable.

Conclusions and Future Direction

Conclusions

The goal of the experiment was to test whether the backward and graded associations that have been extensively documented across a variety of episodic recall paradigms, would also be present in a statistical learning paradigm. From the probabilistic SRT data, there is evidence that both backward and forward-graded associations are formed in a probabilistic SRT task. Moreover, results from an explicit variant of the probabilistic SRT task, the interposed generation task, provided evidence that Ss were not able to explicitly generate the probabilistically most predicted elements, despite displaying sensitivity to these statistics. This data, taken in conjunction with evidence from the episodic recall literature, provides support to the hypothesis that, given a structured input environment, backward and forward-graded associations may automatically be formed as a general property of memory, regardless of the specific application.

Future Directions

Future work that could build upon the present findings to provide converging evidence for backward and forward-graded associations in the context of statistical learning tasks could involve varying the stimulus material, diversifying the S population, and the manipulating the experimental design.

First, the stimulus material of the experiment could be varied, and the experiment replicated using non-verbal material. For example, sequences of visually presented shape-elements, rather than letter-elements, could be employed to test the domain generality of the backward and forward-graded associations that were demonstrated in this verbal variant of the task.

Second, while younger adults comprised the S population of the present experiment, older adults are also a population of interest. Older adults demonstrate an associative memory deficit that has been attributed to difficulty binding item information to contextual information (e.g., Naveh-Benjamin, 2000; Provyn et al., 2007). A deficit in the ability to bind sequence elements to the context of the underlying structure from which the elements are drawn might be manifest as attenuated backward and forward-graded associations. Therefore, if item-to-context binding is an associative process that gives rise to the backward and forward-graded asssociations in the present experiment, and if older adults have impaired item-to-context binding abilities, older adults may not demonstrate the marked backward and forwad-graded associations present in the young adults' RT data.

Finally, the underlying structure from which the sequence elements are drawn could be manipulated. While the underlying ring structure was selected for its robust simplicity, this design does not necessarily reflect the complexity of natural languge. Statistical learning is often used to examine natural language on a smaller, more controlled scale. Therefore, the case to be made for these lag-effects could be strengthened by converging evidence from other statistical learning designs. A design of concentric rings, with conditional probabilities for lags both within and across rings could be employed. Lag could also be measured in artificial grammar tasks. While artificial grammars are arguably more complex than a simple ring, grammars could be designed to examine category rules, induction and violations (see Hunt & Aslin, 2010) as an extension of the more basic concept of lag.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by human infants. *Psychological Science*, *9*, 321-324.

Bunsey, M., & Eichenbaum, H. B. (1996). Conservation of hippocampal memory function in rats and humans. *Nature*, *379*(6562), 255-257.

Canfield, R. L., & Haith, M. M. (1991). Young infants' visual expectations for symmetric and asymmetric stimulus sequences. *Developmental Psychology*, *27*, 198-208.

Chapman, L. J., Chapman, J. P., Curran, T., & Miller, M. B. (1994). Do children and the elderly show heightened semantic priming? how to answer the question. *Developmental Review*, *14*, 159-185.

Cleeremans, A. (1993). *Mechanisms of implicit learning: Connectionist models of sequence processing.* Cambridge, MA, USA: The MIT Press.

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

Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure of sequence events. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *16*, 17-30.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159.

Cole, R., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (p. 133-163). Hillsdale, NJ: Erlbaum.

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: pitting abstract against stimulus-specific representations. *Psychological Science*, *17*(10), 905-12.

Conway, C. M., Goldstone, R. L., & Christiansen, M. H. (2007). Spatial contraints on statistical learning of multi-element scenes. In D. S. McNamara & J. G. Tafton (Eds.), *Proceedings of the 29th annual conference of the cognitive science society* (p. 185-190). Austin, TX: Cognitive Science Society.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of non-adjacent dependencies in tone sequences. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *30*, 1119-1130.

Cunillera, T., Camara, E., Laine, M., & Rodriguez-Fornells, A. (2010). Words as anchors: Known words facilitate statistical learning. *Experimental Psychology*, *57*(2), 134-141.

Curran, T. (1997). Higher-order associative learning in amnesia: Evidence from the serial reaction time task. *Journal of Cognitive Neuroscience*, *9*, 522-533.

Curran, T., & Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *19*, 189-202.

Destrebecqz, A., & Cleermans, A. (2001). Can sequence learning be implicit? new evidence with teh process dissociation procedure. *Psychonomica Bulletin and Review*, *8*(2), 343-350.

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 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*(3), 458-67.

Frensch, P. A., & Miner, C. S. (1994). Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory and Cognition*, 95-110.

Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J. Y., Schwade, J. A., Onnis, L., & Edelman, S. (2010). General cognitive principles for learning structure in time and space. *Trends in Cognitive Science*, *14*, 249-258.

Gomez, R. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431-436.

Harris, Z. S. (1955). From phoneme to morpheme. *Language*, *31*, 190-222.

Howard, M. W., Jing, B., Rao, V. A., Provyn, J. P., & Datey, A. V. (2009). Bridging the gap: transitive associations between items presented in similar temporal contexts. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *35*(2), 391-407.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *25*, 923-941.

Hunt, R. H., & Aslin, R. N. (2010). Category induction via distributional analysis: Evidence from a serial reaction time task. *Journal of Memory and Language*, *62*(2), 98-112.

Jackson, G., & Jackson, S. (1992). *Sequence structure and sequential learning: The evidence from ageing reconsidered* (Technical Report No. 92-9). Institute of Cognitive Decision Sciences, University of Oregon.

Jackson, G. M., & Jackson, S. R. (1995). Do measures of explicit learning actually measure what is being learnt in the serial reaction time task? *Psyche*, *2*(20).

Jimenez, L., Mendez, C., & Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *22*, 948-969.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*, 1-23.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*, 103-109.

Kahana, M. J., & Howard, M. W. (2005). Spacing and lag effects in free recall of pure lists. *Psychonomic Bulletin & Review.*, *12*, 159-164.

Kahana, M. J., Howard, M. W., & Polyn, S. M. (2008). Associative retrieval processes in episodic memory. In H. L. Roediger (Ed.), *Learning and memory: A comprehensive reference* (p. 476-490). Oxford: Academic Press.

Kahana, M. J., & Loftus, G. (1999). Response time versus accuracy in human memory. In R. J. Sternberg (Ed.), *The nature of cognition* (p. 322-384). Cambridge, MA: MIT Press.

Keele, S. W., Jennings, P., Jones, S., Caulton, D., & Cohen, A. (1995). On the modularity of sequence represenations. *Journal of Motor Behavior*, *27*, 17-30.

Kilic, A., Criss, A. H., & Howard, M. W. (2013). A causal contiguity effect that persists across time scales. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *39*(1), 297-303.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-42.

Klein, K. A., Addis, K. M., & Kahana, M. J. (2005). A comparative analysis of serial and free recall. *Memory & Cognition*, *33*(5), 833-839.

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, *3*, 172-192.

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *20*(1), 79-91.

Kuhn, G., & Dienes, Z. (2005). Implicit learning of nonlocal musical rules: implicitly learning more than chunks. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *31*(6), 1417-1432.

Kuhn, G., & Dienes, Z. (2008). Learning non-local dependencies. *Cognition*, *106*(1), 184-206.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffries (Ed.), *Cerebral mechanisms in behavior* (p. 112-136). New York, NY: Wiley.

Lewicki, P., Hill, T., & Bizot, E. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. *Cognitive Psychology*, *20*, 24-37.

Lew-Williams, C., & Saffran, J. R. (2012). All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition*, *122*, 241-246.

Merikle, P. M., & Reingold, E. M. (1991). Comparing direct (explicit) and indirect (implicit) measures to study unconscious memory. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *17*, 224-233.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, *1*(31), 1-9.

Moon, C., Panneton-Cooper, R., & Fifer, W. (1993). *Infant Behavior and Development*, *16*, 495.

Murdock, B. B., & Okada, R. (1970). Interresponse times in single- trial free recall. *Journal of Verbal Learning and Verbal Behavior*, *86*, 263-267.

Naveh-Benjamin, M. (2000). Adult-age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *26*(5), 1170-1187.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance i. statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127-62.

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*(2), 85-117.

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

Onnis, L., Christiansen, M. H., Chater, N., & Gomez, R. (2003). Reduction of uncertainty in human sequential learning: Preliminary evidence from artificial grammar learning. *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, 886-891.

Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingrediant for detecting and generalizing in nonadjacent dependencies. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, 1047-1052.

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

Perruchet, P., & Amorim, M. A. (1992). Conscious knowledge and changes in performance in sequence learning: evidence against dissociation. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *18*(4), 785-800.

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

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*(4), 573-83.

Popper, J. (1959). Mediated generalization. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory* (p. 94-108). Stanford, CA: Stanford University Press.

Primoff, E. (1938). Backward and forward associations as an organizing act in serial and in paired-associate learning. *Journal of Psychology*, *5*, 375-395.

Provyn, J. P., Sliwinski, M. J., & Howard, M. W. (2007). Effects of age on contextually mediated associations in paired associate learning. *22*(4), 846-57.

Reber, P. J., & Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learning and Memory*, *1*(4), 217-29.

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learnt about sequence structure. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *20*, 585-594.

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

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

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*(1), 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.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*, 101-105.

Schendan, H. E., Searl, M. M., Melrose, R. J., & Stern, C. E. (2003). An fmri study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron*, *37*(6), 1013-25.

Schvaneveldt, R. W., & Gomez, R. L. (1998). Attention and probabilistic sequence learning. *Psychological Research*, *61*, 175-190.

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, *115*(4), 893-912.

Sederberg, P. B., Miller, J. F., Howard, M. W., & Kahana, M. J. (2010). The temporal contiguity effects predicts episodic memory performance. *Memory & Cognition*, *38*(6), 689-699.

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *16*, 592-608.

Shanks, D. R., & Johnstone, T. (1999). Evaluating the relationship between explicit and implicit knowledge in a sequential reaction time task. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *25*, 1435-1451.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*, 367-447.

Slamecka, N. J. (1976). An analysis of double-function lists. *Memory & Cognition*, *4*, 581-585.

Stadler, M. A. (1992). Statistical structure and implicit serial learning. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *18*, 318-327.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*, 706-716.

Tichner, E. B. (1909). *Lectures on the experimental psychology of the thought processes.* New York, NY: Macmillan.

Tomblin, J. B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural learning in adolescents with and without specific language impairments. *Language, Learning and Development*, *3*, 269-293.

Turk-Browne, N. B., Junge, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*, 552-564.

Umemoto, T., & Hilgard, E. R. (1961). Paired-associate learning as a function of similarity: common stimulus and response items within the list. *Journal of Experimental Psychology*, *62*, 97-104.

Vapnik, V. N. (1998). *Statistical learning theory.* New York, NY: Wiley-Interscience.

Vouloumanos, A., Brosseau-Liard, P. E., Balaban, E., & Hager, A. (2012). Are the products of statistical learning abstract or stimulus-specific? *Frontiers in Psychology*, *3*(70), 1-11.

Willingham, D. B. (1999). Implicit motor sequence learning is not purely perceptual. *Memory & Cognition*, *27*(3), 561-572.

Willingham, D. B., Greenley, D. B., & Bardone, A. M. (1993). Dissociation in a serial response time task using a recognition measure: Comment on perruchet and amorim. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *19*, 1424-1430.

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *15*(6), 1047-60.

Willits, J. A., Seidenberg, M. S., & Saffran, J. R. (2009). Verbs are looking good in early language acquisition. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 2570-2575.

Young, R. K. (1961). Paired-associate learning when the same items occur as stimuli and responses. *Journal of Experimental Psychology*, *61*, 315-8.

Zaromb, F. M., Howard, M. W., Dolan, E. D., Sirotin, Y. B., Tully, M., Wingfield, A., & Kahana, M. J. (2005). Temporally-based false memories in free recall. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *32*(4), 792-804.

Curriculum Vitae

Jennifer Provyn
jpprovyn AT syr DOT edu
jennifer DOT provyn AT quintiles DOT com
Quintiles, Inc.
6700 West 115th Street
Overland Park, KS 66211
913.708.6983 (P)
913.708.6212 (F)

Personal

Born 9.28.80, Overland Park, KS
Citizenship: USA

Professional Positions

- 2012-present, Investigator Contract Specialist, Investigator Contract Services, Quintiles, Inc.
- 2011-2012, Associate Regulatory and Start-Up Specialist, Regulatory and Start-Up, Quintiles, Inc.
- 2009-2010, Research and Technology Associate, Center for Research and Evaluation, Edvantia, Inc.

Education

- Ph.D 2013 Syracuse University, Experimental Psychology; Dissertation: *Associative processes in statistical learning: Paradoxical prediction of the past*
- M.S. 2007 Syracuse University, Experimental Psychology; Thesis: *Effects of age on contextually mediated associations in paired-associate learning*
- B.A. 2003 Rockhurst University, Magna Cum Laude, Psychology
- B.A. 2003 Rockhurst University, Magna Cum Laude, Philosophy

Peer-reviewed Publications

- Provyn, J. P., Sliwinski, M. J. and Howard, M. W. (2007). Effects of age on contextually mediated associations in paired-associate learning. *Psychology & Aging,* **22**, 846-857.
- Howard, M. W., Jing, B. Rao, V. A., Provyn, J. P. and Datey, A. V. (2009). Bridging the gap: Transitive associations between items presented in similar temporal contexts. *Journal of Experimental Psychology: Learning, Memory and Cognition,* **35**, 391-407.

Working Papers

- Provyn, J. P. and Howard, M.W. (in preparation). Associative processes in statistical learning: Paradoxical prediction of the

past

Support/Awards

- 2007-2008 Eric F. Gardner Fellowship
- 2006 Syracuse University one-month summer fellowship: $700.00
- 2005 Syracuse University one-month summer fellowship: $700.00

Invited Talks

- “Effects of age on contextually mediated associations in paired-associate learning, Garnder Conference, September 2007, Auburn, NY.

Teaching and Mentoring

- Teaching Assistant: PSY 205, Foundations of Human Behavior, 2003-2005
- Instructor: PSY 332, Cognitive Psychology Lab, 2006-2009
- Guest lecturer for PSY 373 (Human Memory), Fall 2006
- Guest lecturer for PSY 373 (Human Memory), Spring 2008
- Teaching Assistant: PSY 323 Cognitive Psychology, 2007-2008
- Teaching Assistant: PSY 205, Foundations of Human Behavior, Summer Session II, 2008
- Psychology Research Coordinator: Summer Session II, 2008
- Tutor: PSY 252, Statistical Methods II, 2008-2009
- Adjunct Psychology Professor, Donnelly College, 2011

Presentations at Scientific Meetings

- Provyn, J. P. and Howard, M. W. Associative processes in probabilistic sequence learning. *Psychonomic Society*, November 2007, Long Beach, CA.
- Provyn, J. P., Sliwinski, M. J. and Howard, M. W. Age differences in transitive associations, Poster presented to the *Society of Mathematical Psychology*, Vancouver, BC, July 2006.