Item-to-Item Associations Contribute to Memory for Serial Order

By

Dakota R. B. Lindsey

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

August 9th, 2019

Nashville, Tennessee

Approved:

Gordon D. Logan, Ph.D.

Lisa K. Fazio, Ph.D.

Sean M. Polyn, Ph.D.

Duane G. Watson, Ph.D.

Geoffrey F. Woodman, Ph.D.

To Emily and Rhys, who make every day a joy

ACKNOWLEDGEMENTS

I want to extend a special thanks to the following people:

Gordon, for being a great advisor, colleague, and friend;

My dissertation committee members, Duane, Geoff, Lisa, and Sean, for helping me develop the ideas and methods in this project;

My peers, particularly Jana, Jeff, and Jon, for miscellaneous help throughout this project, and for all the fun and interesting conversations over the years;

Randy, for getting me interested in psychology, and for giving me the skills needed to get into graduate school;

My family and friends, particularly Emily, Rhys, Aaron, Tyler, and Yesh, for all the personal support they have given me through the years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

IT	Initiation time
IRT	Inter-response time
BF	Bayes factor
df	Degrees of freedom
MSE	Mean squared error
a	Estimate of peak performance, given by fitting a power function to data
b	Estimate of the amount that performance can improve over practice, given by fitting power function to data
c	Estimate of the rate of learning, given by fitting a power function to data
RMSD	Root mean squared deviation
S.E.	Standard error

CHAPTER 1

INTRODUCTION

Nearly every task we face, ranging from the simplest and most routine to the most complex and novel, must be completed in a specific order. Calling a friend to chat requires recalling the digits of his or her phone number in the correct order. Speaking to this friend requires you to articulate sounds in a particular order. We often manage to produce these sequences without error, even though there are billions of phone numbers that could be dialed and an endless number of utterances that could be spoken. The problem of serial order, first formally described by Lashley (1951), has been a topic of interest in psychology for nearly 70 years. However, despite decades of research on the topic, the mechanism of serial order is still hotly debated.

Much of our current understanding of the serial order mechanism stems from research in the domain of serial memory. Research in this domain is most often done using the serial recall task. Recalling your friend's phone number is essentially a serial recall task: The goal is to recall a sequence of items in their appropriate order. Most often, the stimuli used in serial recall tasks are novel sequences of letters or words. These novel sequences are presented to the participant, and the participant must reproduce the sequences after some delay (typically immediately after presentation). The focus in this task is how often errors occur, which items are involved in an error, and, less frequently, the timing of responses. These data are used to inform the mechanism that produces serial order.

The earliest explanations of serial recall performance are associative chaining theories, traceable back to Ebbinghaus (1885). In associative chaining theories, the items in a sequence are associated with one another. Once retrieved, each item then serves as a retrieval cue to recall from memory the item that followed it in the sequence. These theories provide an intuitive explanation of serial memory retrieval, but they predict that people will struggle to recover from error: Retrieval of an incorrect item produces an incorrect retrieval cue, so the next item recalled is likely to also be erroneous. In reality, people recover from error more often than not. Making some of the items in the sequence similar to one another (e.g., by making items phonologically or visually similar; Henson, Norris, Page, & Baddeley, 1996; Logie et al. 2016) impairs retrieval

of the similar items, but it does not impair retrieval of the items that follow the similar items in the sequence. Under "typical" serial recall conditions (i.e., short lists of items selected from a very limited set), people show a tendency to move backward in the sequence to report an item after skipping it, in effect correcting their mistake (Farrell, Hurlstone, & Lewandowsky, 2013; Osth & Dennis, 2015).

These findings, along with a growing body of evidence suggesting that people code items with respect to their position in a sequence (Ebenholtz, 1963; Fischer-Baum & McCloskey, 2015; Henson, 1996; Henson, Norris, Page, & Baddeley, 1996; Lewandowsky & Farrell, 2008; etc.), led to the demise of associative chaining theories. Positional theories are now the most common explanation of serial order (for review, see Hurlstone, Hitch, & Baddeley, 2014; Lewandowsky & Farrell, 2008). These theories posit that each item is associated with a representation of its position in the sequence (e.g., the temporal context at the time the item was presented; Brown, Preece, & Hulme, 2000) and that retrieval of the sequence is accomplished by stepping through the set of positional representations, retrieving the item associated with each.

Nearly all positional theories eschew associations between items, effectively assuming that item-to-item associations play no role in memory for serial order (but see Burgess & Hitch, 1992). The lack of impairment in performance following phonologically similar items has been used as evidence for this assumption: If an erroneous item has any influence on retrieval of the next item, then retrieval of the next item should be impaired. However, this argument implicitly assumes that the phonological representations of items are the retrieval cues. In contrast, some positional theories (e.g., Henson, 1998; Page & Norris, 1998) assume that items are represented separately from their phonology. First a categorical representation of an item is retrieved, and then this representation retrieves its associated phonology. If the categorical representations of items are the retrieval cues instead of the phonological representations, then mistakes made during phonological selection would not feed forward into retrieval. Consequently, accuracy would be no worse after a phonologically confusable item (cf. Caplan, 2015). The rejection of associative chaining theories, which rely solely on item-to-item associations, is well-founded: There is strong evidence for some role of position-to-item associations. However, there is no conclusive evidence supporting the assumption that item-to-item associations play no role at all in serial recall.

Outside of typical serial recall conditions, specifically when the length of the to-be-recalled sequences is long, people tend to continue forward in a sequence after skipping an item, resulting in the skipped item being omitted entirely (Solway, Murdock, & Kahana, 2012). This pattern is consistent with retrieval using item-to-item associations because these associations are stronger in the forward direction than in the backward direction (Raskin & Cook, 1937). When using an item as a retrieval cue, the item that is most likely to be recalled next is the one that followed the cue in the sequence. This finding, however, does not conclusively favor retrieval using item-to-item associations. Farrell (2012; also Farrell, Hurlstone, & Lewandowsky, 2013) demonstrated that the tendency to move forward could be produced in the absence of item-to-item associations if people subjectively organize a sequence into groups.

At present, there is no strong evidence for or against the use of item-to-item associations in serial recall. The goal of the research presented here is to obtain more conclusive evidence, using a paradigm that more sharply tests for item-to-item associations than previous research.

The Spin List Procedure

The spin list procedure is a promising avenue for testing item-to-item associations. In its original incarnation, devised by Ebenholtz (1963), people practiced recalling two types of sequence: those in which the same items were presented in the same serial position on each repetition, and those in which the same items were presented but the items were "spun" to different positions (e.g., ABCDEF → FABCDE). In identical sequences but not spun sequences, the serial position of each item remained consistent across repetitions. Identical lists were learned more quickly than spun sequences, so learning was sensitive to consistencies in the serial positions of items in the sequence. This result suggests that people learn associations between items and their serial positions, and that serial position can be used as (or at least contributes to) a retrieval cue.

Kahana, Mollison, and Addis (2010) replicated and reevaluated Ebenholtz's (1963) findings: Even though learning was slower for spun sequences due to inconsistent serial position, learning did occur in these sequences. They argued that the learning observed in spun sequences is consistent with retrieval using item-to-item associations. When relative order is consistent over practice, associations between items should strengthen. If retrieval makes use of these

associations (i.e., if items are used as cues to retrieve other items), then performance should improve over practice even when serial position is inconsistent.

However, there are other explanations for this learning. Kahana, Mollison, and Addis (2010) admit that learning in spun sequences would be expected if people "learn the circular nature of the lists and develop a method by which they can map the items into a circular coordinate scheme" (p. 99). Conscious detection of the structure of spun lists may lead to strategies, such as the one Kahana, Mollison, and Addis (2010) describe, that improve the memorability of these sequences. Spun sequences are also "closed sets" by nature, each composed of the same small set of items. When the set of items is limited in this manner, it is easier to remember each individual item and to guess their positions in the sequence, compared to when an unrestricted set is used (Osth & Dennis, 2015). It is important to demonstrate that learning can occur in spun sequences without conscious detection of the spin, and that the magnitude of learning is greater than what would be expected by having a small repeating set of letters.

Lindsey and Logan (2019) accomplished this in a set of experiments aimed to test for item-to-item associations in typing. They had participants practice typing two types of sequence: spun sequences, and scrambled sequences in which the order of letters was set according to a balanced Latin square. A balanced Latin square ensured that each letter appeared in each serial position, and each letter appeared before and after each other letter. As a result, both serial position and relative position were inconsistent in scrambled sequences over practice. Learning was observed for both types of sequence, possibly because both used closed sets of letters. However, learning was greater for spun sequences, suggesting that the learning observed in spun sequences was not solely attributable to the restricted letter sets. A post-experiment interview revealed that few of the participants detected any additional structure in spun sequences, so the advantage for spun sequences could not have reflected strategy use. They argued that the only viable explanation of their results was that typing makes use of item-to-item associations.

Lindsey and Logan (2019) provided strong evidence for the use of item-to-item associations in typing, but this may not generalize to serial recall. In tasks like serial recall, items are stored in working memory, so people presumably have conscious, top-down control over the order in which the items are reported. In typing, on the other hand, people are largely unaware of the keys they press (Logan & Crump, 2010) or where their fingers are moving on the keyboard

(Snyder et al., 2014). The manner in which serial order is controlled consciously through working memory need not be the same manner in which it is controlled unconsciously in the motor system.

The research presented in this paper adapts the methods of Lindsey and Logan (2019) to the domain of serial memory. The general method is nearly identical: Participants see sequences of letters, presented simultaneously, and must type the sequences in their correct (left to right) order. In their experiments, however, each sequence remained on screen while it was typed, so the task did not require memorization of the sequence. In the proposed experiments, the sequences are presented briefly before disappearing, and therefore must be memorized. Because they must be memorized, errors made within trials and performance changes between trials should reflect failures and learning adjustments made by the serial memory system.

CHAPTER 2

TESTING FOR ITEM-TO-ITEM ASSOCIATIONS

In this chapter, I present two experiments designed to address the primary question of the paper: whether or not item-to-item associations are used in serial recall. In these experiments, I compare performance on spun lists to performance on Latin square scrambled lists. The only structural difference between these list types is relative order: in spun lists it is consistent, and in scrambled lists it is not. This difference is key – consistent relative order allows item-to-item associations to strengthen. If these associations are used in serial recall, then spun lists should be easier to remember than scrambled lists. These associations should transfer, making it easier to remember new lists in which the relative order of items is consistent with the order that was learned before.

Experiment 1

Experiment 1 aims to establish the foundational effect of the paper: better learning for spun lists. If item-to-item associations are used in serial recall, then performance on spun lists should improve more quickly than performance on scrambled lists due to the consistent relative order of the items. If spun lists are not learned any more quickly than scrambled lists, then memory for serial order does no use item-to-item associations.

Method

Participants. Consistent with Lindsey and Logan (2019) and other studies examining practice effects on the typing of non-word lists (e.g., Yamaguchi & Logan, 2016), I tested 24 participants. The participants were native English speakers between the ages of 18 and 35 who reported normal or corrected-to-normal vision and reported a typing speed greater than 40 words per minute. The participants were tested in 1.5 hour timeslots and received $18 or course credit as compensation.

Apparatus and stimuli. I used E-prime 2.0 (Psychology Software Tools) to present stimuli, record responses, and record the accuracy and timing of responses. The task was administered on ASUS M32BF desktop computers with BenQ XL2411Z flat screen monitors. Responses were taken from standard QWERTY keyboards with rubber dome switches. Only the letter keys and spacebar were enabled for the task. This apparatus was used for all of the present experiments.

I generated four non-overlapping sets of six letters by sampling randomly from the alphabet without replacement. The letters 'a' and 'e' were excluded from selection to reduce the chance of producing word-like lists. I generated six lists for each of the four sets of letters, ensuring that each letter occupied each position in the list once.

For two of the letter sets, I assigned each of the six letters to a position and then "spun" them to produce the spun lists (Ebenholtz, 1963; Kahana, Mollison, & Addis, 2010). Imagine that the letters occupy positions along a circle (e.g., picture a clock face) – a list is created by reading the letters off of this circle starting with the top position and proceeding clockwise (see Figure 1). A clockwise spin shifts each letter one position clockwise along the circle. Reading the letters off the circle after this clockwise spin produces a list with letters that have shifted rightward one position – except the letter at the end, which shifts leftward 5 positions to the start of the new list. Spinning the list five times ensures that each letter occupies each position, but each letter follows and is followed by only one other letter. The serial positions of the letters are inconsistent, but the relative positions of the letters are consistent. For the other two letter sets, I inserted the letters into a balanced Latin square to produce scrambled lists. The balanced Latin square ensured that there was no consistency in the serial positions or the relative positions of the letters in the scrambled lists.

Participants practiced 24 different 6-letter non-word lists over the course of the experiment. Twelve of these lists – six from one of the spun list sets and six from one of the scrambled list sets – were practiced in the first half of the experiment, and the other twelve were practiced in the second half of the experiment. Table 1 shows example lists that a participant could have seen in this experiment.

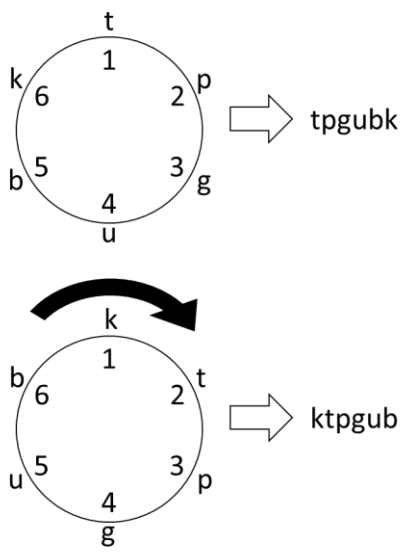

Figure 1. Pictorial depiction of the construction of spun lists. Letters are represented on a circular dial. To create a list, letters are read off of the dial starting with the top letter and proceeding clockwise. Doing this for the first dial produces the list "tpgubk." To create a "spin" of the first list, each letter is rotated one position clockwise along the dial. The letter 'k' is now the starting point, so reading the letters off of the dial now produces the list "ktpgub." For a set of N items, there are N potential starting points, and thus N potential spins of the set.

Table 1
Construction of letter sequences in Experiment 1.

Spun Set								Scrambled Set						
First Half								First Half						
	Item 1	t	p	g	u	b	k		Item 7	q	f	w	m	s	l
	Item 2	p	g	u	b	k	t		Item 8	f	m	q	l	w	s
	Item 3	g	u	b	k	t	p		Item 9	m	l	f	s	q	w
	Item 4	u	b	k	t	p	g		Item 10	l	s	m	w	f	q
	Item 5	b	k	t	p	g	u		Item 11	s	w	l	q	m	f
	Item 6	k	t	p	g	u	b		Item 12	w	q	s	f	l	m
Second Half								Second Half						
	Item 1	y	z	c	n	r	d		Item 7	x	h	o	i	j	v
	Item 2	z	c	n	r	d	y		Item 8	h	i	x	v	o	j
	Item 3	c	n	r	d	y	z		Item 9	i	v	h	j	x	o
	Item 4	n	r	d	y	z	c		Item 10	v	j	i	o	h	x
	Item 5	r	d	y	z	c	n		Item 11	j	o	v	x	i	h
	Item 6	d	y	z	c	n	r		Item 12	o	x	j	h	v	i

These letters serve as examples. Each subject received a different random selection of letters. When presented on the computer screen, the spaces between letters were removed.

Procedure. Before beginning the task, participants were informed that a list of letters would briefly be presented on the screen, that they should memorize the list because they would be asked to type it, that they should type as quickly and accurately as possible, that they should guess if they could not remember the letter that occupied a position, and that mistakes could not be corrected because the backspace key was disabled. Participants were not told about the two list types.

At the beginning of each trial, a fixation cross was presented in the center of the screen for 500 ms. The fixation cross disappeared and was replaced by the list of letters, which remained on screen for 500 ms. A 500 ms blank screen separated the end of the list presentation and the response screen. The text "Response:," presented in the center of the screen, cued participants to type the letters they could remember from the most recently presented list. Letter keys that they pressed were echoed on the screen under the response cue, so participants could see their prior responses. Participants submitted their responses by pressing the spacebar, which cleared the screen and initiated the next trial. A 1 sec blank screen separated the spacebar press and the fixation cross on the next trial. This same trial procedure is used in each experiment.

In each half of the experiment, participants saw only 12 lists (6 spun and 6 scrambled), and each was presented 40 times. Participants completed a total or 960 trials – 480 in each half of the experiment. Presentations were blocked by number of repetitions: No list could be presented again until all of the other lists had been presented. Within a repetition block, the twelve lists were selected randomly without replacement. Self-paced breaks were allowed every 240 trials (roughly every 20 minutes), during which participants could get water, use the restroom, or walk around the lab. Participants were not explicitly informed about the switch between letter sets in the first and second halves of the experiment (though, the halves were separated by a break). After finishing the experiment, I asked participants what they thought of the task, what strategies they used to remember the letters, and whether they noticed any patterns in the order that the letters appeared.

Results and Discussion

In this experiment and others that follow, I am interested in three measures of performance: error rate (defined by the proportion of trials in which at least one letter is recalled erroneously), initiation time (IT; the latency from the onset of the response cue to the first letter

response), and inter-response time (IRT; the average latency between letter responses). With practice on a list, I expect each of these performance measures to improve, but they may improve for different reasons. Improvements in error rate may reflect the accuracy of encoding items from the display, retrieving items or chunks of items from memory, or executing keystrokes. Improvements in IT partially reflect the time to retrieve the first item, but they also may reflect the time to encode the response cue, prepare to respond, or retrieve and unpack a chunk (if the items are chunked over practice). Improvements in IRT reflect the time to retrieve items and execute keystrokes, but also reflect the time to retrieve and unpack chunks (if the list is represented as two or more chunks). As is common with recall of novel lists, the primary focus is on error rate. Although participants were instructed to respond as quickly and accurately as possible, some will inevitably choose to respond at their own pace, making the timing measures less informative indicators of learning.

I calculated average error rates, ITs, and IRTs separately for each participant (24), list type (2), and block (40) in the experiment. ITs and IRTs greater than 3 sec were excluded from the averages. I averaged performance in the two halves of the experiment because they were identical in structure (e.g., block 1 spun list IT in the first half was averaged with block 1 spun list IT in the second half).

My primary interest in this experiment is whether error rate, IT, or IRT improve more for spun sequences than for scrambled sequences over practice. In the current experiments, I tested for this difference in two ways. First, I tested the difference with an ANOVA list type main effect. This test would be sensitive to pre-experimental differences in performance on the two types of list. However, I had no a priori reason to expect a pre-experimental difference; the stimuli were nonsensical letter lists that participants presumably had never seen before, and letters were randomly assigned to the two list type conditions. A t-test comparing first block performance on the two list types was conducted to check for any “accidents” of randomization (it does not ensure equal difficulty for the two list types). A null finding is meaningful for this t-test, so the Jeffrey-Zellner-Siow Bayes factor (BF; Rouder et al., 2009) is reported to provide additional evidence for the null or alternative hypothesis. The second method for testing the difference was through fitting power functions to the data and checking for differences in power function parameters.

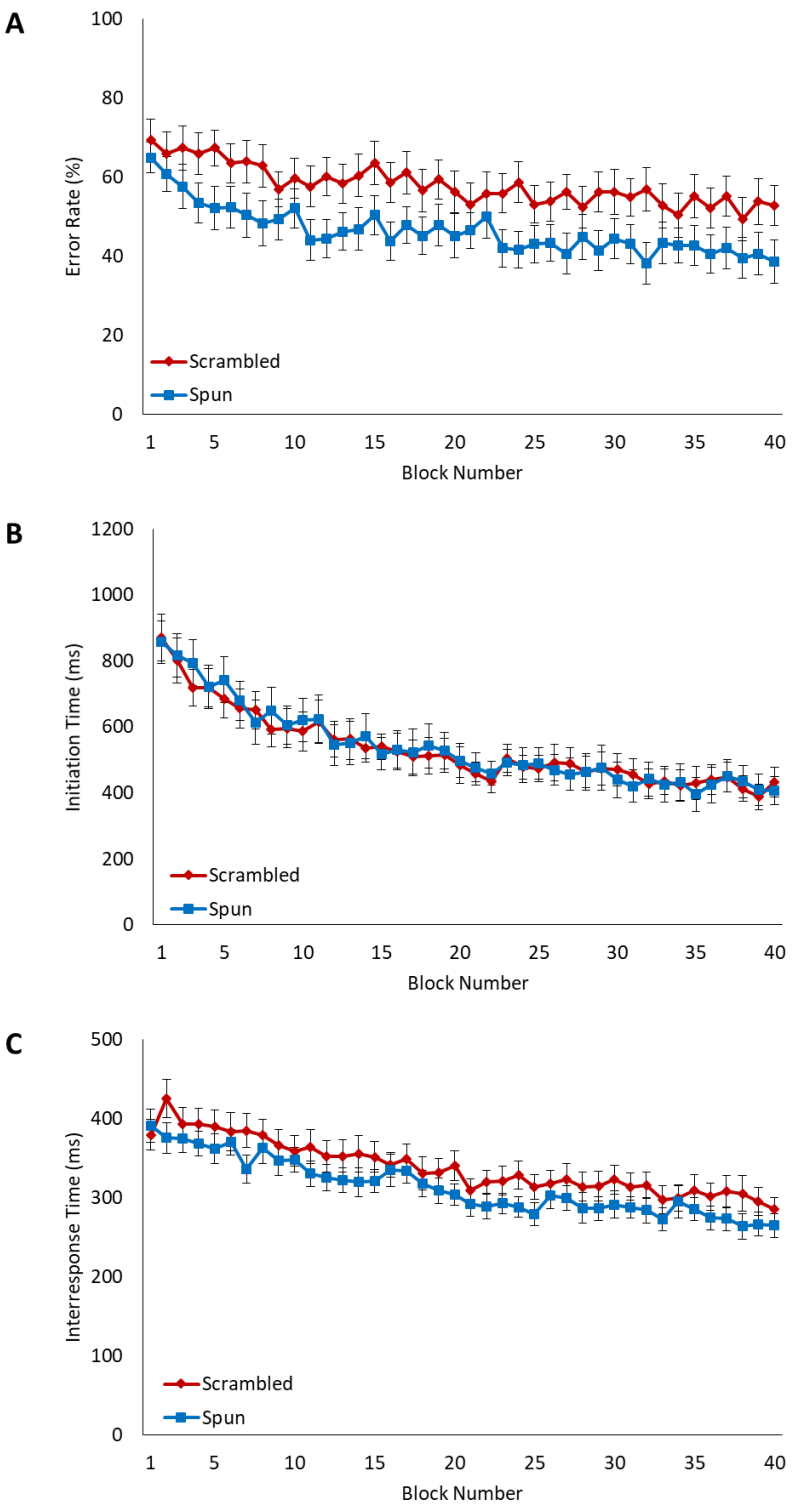

Figure 2. Experiment 1: Mean error rate (Panel A), IT (Panel B), and IRT (Panel C) for each list type as a function of the presentation number of the list (block number). The bars are standard errors of the means.

Mean error rate, IT, and IRT in each block are displayed in Figure 2. Mean performance was analyzed using 2 (list type) x 40 (block) ANOVA. The main effect of list type and the simple main effect of block number in each list type are presented in Table 2. Participants recalled letters from both scrambled and spun lists more accurately and more quickly with practice, supported by significant simple main effect analyses of presentation number on error rate and IRT for each list type (Table 2). Participants also initiated retrieval of each list type more quickly with practice, supported by simple main effect analyses of presentation number on IT (Table 2).

The learning in scrambled lists highlights the importance of using these lists for a baseline comparison, and it undermines the conclusion of Kahana, Mollison, and Addis (2010). Learning occurred in the absence of consistent serial position or relative position, so the learning observed in spun lists cannot be attributed item-to-item associations alone. The use of closed letter sets might have contributed to learning. Learning might also have been supported by the development of sequence-level representations, which allow the lists to be individuated and protect them from interference that arises from other lists that share the same letters (cf. Hitch, Fastame, & Flude, 2005). It is critical to show that *more* learning occurred in spun lists than in scrambled lists. The presence of consistent relative order is the only characteristic that separates spun lists from scrambled lists, so any additional learning is attributable to consistent relative order.

Table 2

Experiment 1: ANOVA and simple main effect analyses for training effects.

DV	F	dfs	MSE	p	ε	η_p^2
			Spun vs. Scrambled Performance			
Error Rate	8.254	1, 23	0.795	0.009	1.000	0.264
IT	0.013	1, 23	1,037,908	0.909	1.000	0.001
IRT	2.184	1, 23	140,065	0.153	1.000	0.087
			Learning (Spun)			
Error Rate	6.617	39, 897	0.012	< 0.001	0.478	0.223
IT	29.198	39, 897	11,820	< 0.001	0.273	0.559
IRT	16.000	39, 897	1,889	< 0.001	0.617	0.410
			Learning (Scrambled)			
Error Rate	4.762	39, 897	0.012	< 0.001	0.496	0.172
IT	25.068	39, 897	11,820	< 0.001	0.322	0.522
IRT	14.416	39, 897	1,889	< 0.001	0.203	0.385

Learning analyses are simple main effects analyses of presentation number that use the MSE of the List Type X Presentation interaction. The simple main effect analyses are not corrected for violations of sphericity, but Huynh-Feldt epsilon values are provided to gauge the severity of these violations.

Participants learned the spun lists more quickly than the scrambled lists: Error rates were lower for spun lists than for scrambled lists (46.3% vs. 58.0%), supported by a significant main effect of list type (Table 2). These learning differences do not reflect initial differences in performance because first block performance was not significantly different between the list types (spun: 64.9%; scrambled: 69.4%; Table 3). The learning advantage for spun lists could reflect improvements in item retrieval that arise from the consistent relative order of the items. Thus, the learning advantage observed here is consistent with item-to-item associations being used in serial recall.

ITs and IRTs were no different between the list types over practice (Table 2), suggesting that consistent relative order does not improve the speed of response preparation (spun: 537.1 ms; scrambled: 531.7 ms) or the speed of retrieving items from memory (spun: 313.2 ms; scrambled: 338.4 ms). The lack of significant timing differences may be the result of participants adopting a similar response rhythm for each list.

Table 3

Experiment 1: T-tests for first block performance and power function learning rate.

DV	t	df	M_{A-B}	SE_{A-B}	p	d	BF
Spun vs. Scrambled First Block Performance							
Error Rate	-0.878	23	-0.045	0.252	0.389	-0.199	3.290 (N)
IT	-0.300	23	-14.392	47.975	0.767	-0.043	4.471 (N)
IRT	0.726	23	11.910	16.404	0.475	0.121	3.669 (N)
Spun vs. Scrambled Learning Rate							
Error Rate	2.648	23	0.092	0.035	0.007	0.664	3.576 (A)
IT	-0.400	23	-0.031	0.076	0.346	-0.114	4.330 (N)
IRT	1.168	23	0.088	0.075	0.127	0.350	2.538 (N)

Learning rate analyses are conducted on the c parameters obtained from power function fits. For Bayes factors, numbers followed by (A) indicate evidence in favor of the alternative hypothesis, and numbers followed by (N) indicate evidence in favor of the null hypothesis.

In the error rate graph, the line for spun lists diverges from the line for scrambled lists early in practice, and the difference between the lines remains fairly constant for the remainder of the experiment. This pattern may be indicative of item-to-item associations that are learned with a closed-loop learning rule (cf. Lewandowsky & Murdock, 1989; Solway, Murdock, & Kahana, 2012): there is some maximum amount of association that two items can share; the amount of learning on each pairing of the two items is proportional to the amount of associative strength than can still be gained; and, in this case, the proportion learned on each pairing is high, so the maximum amount of association is approached early in practice.

The post-experiment interviews revealed that only 8 of the 24 participants (33.3%) noticed the consistent relative order in spun lists. Another 9 participants (37.5%) noticed that the spun lists were more structured (e.g., many noticed that groups of three letters would swap places in the spun lists) but did not explicitly tie this structure to consistency in relative order. The remaining 7 participants (29.2%) were unable to detect any structural differences in the lists. I re-ran the error rate ANOVA twice, once with spin detection as an additional factor and once with structure detection as an additional factor. The spun list advantage observed for error rate did not depend on the detection of spin, $F(1, 22) = 2.043$, MSE = 0.761, $p = .167$, nor did it depend on the detection of additional structure in spun lists, $F(1, 22) = 0.344$, MSE = 0.818, $p = .563$. The

advantage for spun lists seems to arise without overt detection of structure, and thus without strategies that exploit that structure.

Like Kahana, Mollison, and Addis (2010), I found evidence that people can use item-to-item associations for serial memory retrieval. The spun list advantage cannot be explained by associations between positional cues and items because serial position is inconsistent for both spun and scrambled lists. Critically, the current experiment also demonstrates that the learning observed in spun lists is not entirely explainable by the use of closed letter sets (spun and scrambled lists were constructed from equally sized letter sets), the development of sequence-level representations (spun and scrambled lists were practiced equally often), or strategy use (although this is based on potentially unreliable survey data and will be revisited in Chapter 4).

Power function fits. Often, the improvement of performance over practice follows a power function, and the parameters of this power function can be theoretically meaningful (cf. Logan, 1988). As an alternative method of analyzing the data, I fit power functions to each participant's data and tested differences in the power function parameters. A major draw to this method is that it reduces the large amount of trial data (40 blocks of trials) into a few meaningful parameters. A major flaw to this method is that parameter estimation is often not perfectly reliable; running the same fitting routine on different occasions can give different parameter estimates.

The functions I fit to the data followed the form:

$$Y = a + bX^{-c}.$$

In the context of the current experiment, Y is a performance measure (error rate, IT, or IRT), X is the experiment block, and a, b, and c are parameters that determine how performance changes as block changes. The a parameter controls peak performance, the b parameter controls how much performance can improve, and together they control the starting point of performance. The c parameter is the learning rate – it controls how quickly performance improves from the starting point to peak performance – with higher values of c reflecting faster rates of learning.

I fit two power functions to each participant's data – one function for each list type. The parameters of these two power functions were estimated simultaneously by minimizing the root mean square deviation (RMSD) between the observed data and the function predicted data. This minimization was accomplished using the fminsearch simplex optimization routine in MATLAB

R2018a (MathWorks). In an attempt to avoid local minima, simplex was run 50 times for each participant using different starting parameter values on each iteration. The parameter estimates were taken from the iteration that produced the largest correlation between the observed and predicted data.

The *a* parameter was fixed to a constant value for all participants. For error rate, *a* was set to 0 because peak performance is making zero errors. For IT and IRT, *a* was set to 150, which corresponds to a minimum keystroke time of 150 ms. This is roughly equivalent to typing 80 words per minute[1], a typing speed that is above average for English words (Crump & Logan, 2013) and thus a reasonable upper limit for the speed of typing novel non-words. The *b* parameter was allowed to differ between participants, but it was constrained to be the same value for spun lists and scrambled lists. Because *a* was fixed and *b* was constrained to be the same for spun and scrambled lists, the two list types were predicted to start at the same level of performance. This forced all of the difference in prediction into the *c* parameter, which was allowed to differ between participants and list types. Both *b* and *c* were constrained to take positive values, and *c* was bounded between 0 and 1. For each participant, 3 parameters were estimated by the optimization routine: a *b* parameter shared between the spun list power function and scrambled list power function, a *c* parameter for the spun list function, and a *c* parameter for the scrambled list function.

The fits were not great, likely due to large amounts of within-subject variability. Average R^2 across the 24 participants was .393 for error rate, .543 for IT, and .500 for IRT. The shared *b* parameter, averaging over the estimates for individual participants, was 72.4 for error rate, 859.7 for IT, and 300.9 for IRT. I conducted a paired-sample t-test on the *c* parameters obtained for the two list types. The *c* parameter represents a participant's learning rate for that a given list type, so a higher *c* parameter for spun lists would indicate faster learning on these lists. The results of these new t-tests matched the results of the previously presented ANOVA: Learning rate was significantly faster for spun lists in terms of error rate (spun: .191; scrambled: .099; Table 3). The spun list learning rate was numerically higher in IRT, but this difference did not reach

[1] Assuming a word length of 5 letters (cf. Logan, Ulrich, & Lindsey, 2016) and equivalent keystroke timing for each letter in the word, the time to type one word would be 150*5 = 750 ms. This is equivalent to one word per 0.75 s, one word per 0.0125 minutes, and 80 words per minute.

significance (spun: .302; scrambled: .215; Table 3). Learning rate was not higher in IT (spun: .364; scrambled: .395; Table 3). The quicker reduction in error rate observed for spun lists is consistent with the use of item-to-item associations.

Experiment 2

The learning advantage for spun lists is consistent with item-to-item associations being learned and used to recall spun lists. If these associations are being used, then, after practicing spun lists, there should also be a performance advantage for unpracticed lists that have the same relative order. Training on the sequence ABCDEF should improve performance on the sequence FABCDE because several of the item-to-item associations carry over (i.e., the associations between A and B, B and C, C and D, D and E). On the other hand, little-to-no positive transfer to unlearned Latin square scrambled sequences would be expected because their relative order is inconsistent.

In Experiment 2, I had participants train on a subset of the 6 lists from each set, and I later tested them on the remaining lists from each set. If item-to-item associations are contributing to recall of spun sequences, then there should be a performance advantage for spun lists in the training portion and in the test portion.

Originally, I had planned two transfer experiments of this nature: one in which participants trained on 4 lists in each set and then tested on the remaining 2, and one in which participants trained on 2 lists in each set and then tested on the remaining 4. I had 24 people participate in each experiment like Experiment 1, but unfortunately both of these transfer experiments were underpowered. Reducing the number of trained lists reduced the number of trials per block and thus the precision in estimating mean performance in each block. Aside from the number of lists on which participants trained and tested, these two experiments were identical in structure. To improve my ability to detect a difference between spun and scrambled lists, I combined the two experiments into one and treated the number of lists manipulation as a grouping variable. Had I anticipated this issue with power, I would have run the number of lists manipulation within-subject instead.

Method

Participants. 48 participants were recruited for this experiment. The 24 participants who trained on 4 lists were recruited before the 24 participants who trained on 2 lists. The selection criteria were the same as Experiment 1. Participants were tested in 1 hour timeslots and received $12 or course credit as compensation.

Apparatus and stimuli. I generated 4 sets of 6 lists in the same manner as Experiment 1. Two sets were spun and 2 sets were scrambled using a balanced Latin square. For half of the participants (24), 4 lists from each set were randomly selected to be training lists, and the remaining 2 lists from each set became test lists. For the other half of the participants (24), 2 lists from each set were randomly selected to be training lists, and the remaining 4 lists from each set became test lists. Table 4 presents an example set of lists.

Procedure. This experiment consisted of a training portion and a test portion. Participants practiced the training lists 40 times each and then practiced the test lists 6 times each. Participants who trained on 4 lists in a set saw 16 different lists during training, 8 different lists during test, and completed 688 trials altogether (640 training + 48 test). Participants who trained on 2 lists in a set saw 8 training lists, 16 test lists, and completed 416 trials (320 training + 96 test). Trials were blocked by repetition, and lists were selected without replacement within each block. Self-paced breaks were offered every 80 trials for all participants. Participants were not explicitly informed that the lists would change during the experiment. All other aspects of the procedure were identical to Experiment 1.

Results and Discussion

Within-subject mean error rate, IT, and IRT were calculated for each participant (48), list type (2) and block (50). The number of lists in a set that were practiced during training did not have a significant impact on performance (Table 5). Because this grouping variable did not interact with any of the other manipulations, I focus on main effects like in Experiment 1. Mean performance for each block, averaged across the participants and group (4 or 2 training lists per letter set), is shown in Figure 3.

Table 4

Construction of letter sequences in Experiment 2.

Spun Set								Scrambled Set						
							Group 1							
Training								Training						
	Item 1	t	p	g	u	b	k		Item 9	q	f	w	m	s	l
	Item 2	p	g	u	b	k	t		Item 10	f	m	q	l	w	s
	Item 3	g	u	b	k	t	p		Item 11	m	l	f	s	q	w
	Item 4	u	b	k	t	p	g		Item 12	l	s	m	w	f	q
	Item 5	y	z	c	n	r	d		Item 13	x	h	o	i	j	v
	Item 6	z	c	n	r	d	y		Item 14	h	i	x	v	o	j
	Item 7	c	n	r	d	y	z		Item 15	i	v	h	j	x	o
	Item 8	n	r	d	y	z	c		Item 16	v	j	i	o	h	x
Test								Test						
	Item 1	b	k	t	p	g	u		Item 5	s	w	l	q	m	f
	Item 2	k	t	p	g	u	b		Item 6	w	q	s	f	l	m
	Item 3	r	d	y	z	c	n		Item 7	j	o	v	x	i	h
	Item 4	d	y	z	c	n	r		Item 8	o	x	j	h	v	i
							Group 2							
Training								Training						
	Item 1	t	p	g	u	b	k		Item 5	q	f	w	m	s	l
	Item 2	p	g	u	b	k	t		Item 6	f	m	q	l	w	s
	Item 3	y	z	c	n	r	d		Item 7	x	h	o	i	j	v
	Item 4	z	c	n	r	d	y		Item 8	h	i	x	v	o	j
Test								Test						
	Item 1	g	u	b	k	t	p		Item 9	m	l	f	s	q	w
	Item 2	u	b	k	t	p	g		Item 10	l	s	m	w	f	q
	Item 3	b	k	t	p	g	u		Item 11	s	w	l	q	m	f
	Item 4	k	t	p	g	u	b		Item 12	w	q	s	f	l	m
	Item 5	c	n	r	d	y	z		Item 13	i	v	h	j	x	o
	Item 6	n	r	d	y	z	c		Item 14	v	j	i	o	h	x
	Item 7	r	d	y	z	c	n		Item 15	j	o	v	x	i	h
	Item 8	d	y	z	c	n	r		Item 16	o	x	j	h	v	i

These letters serve as examples. Each subject received a different random selection of letters. When presented on the computer screen, the spaces between letters were removed.

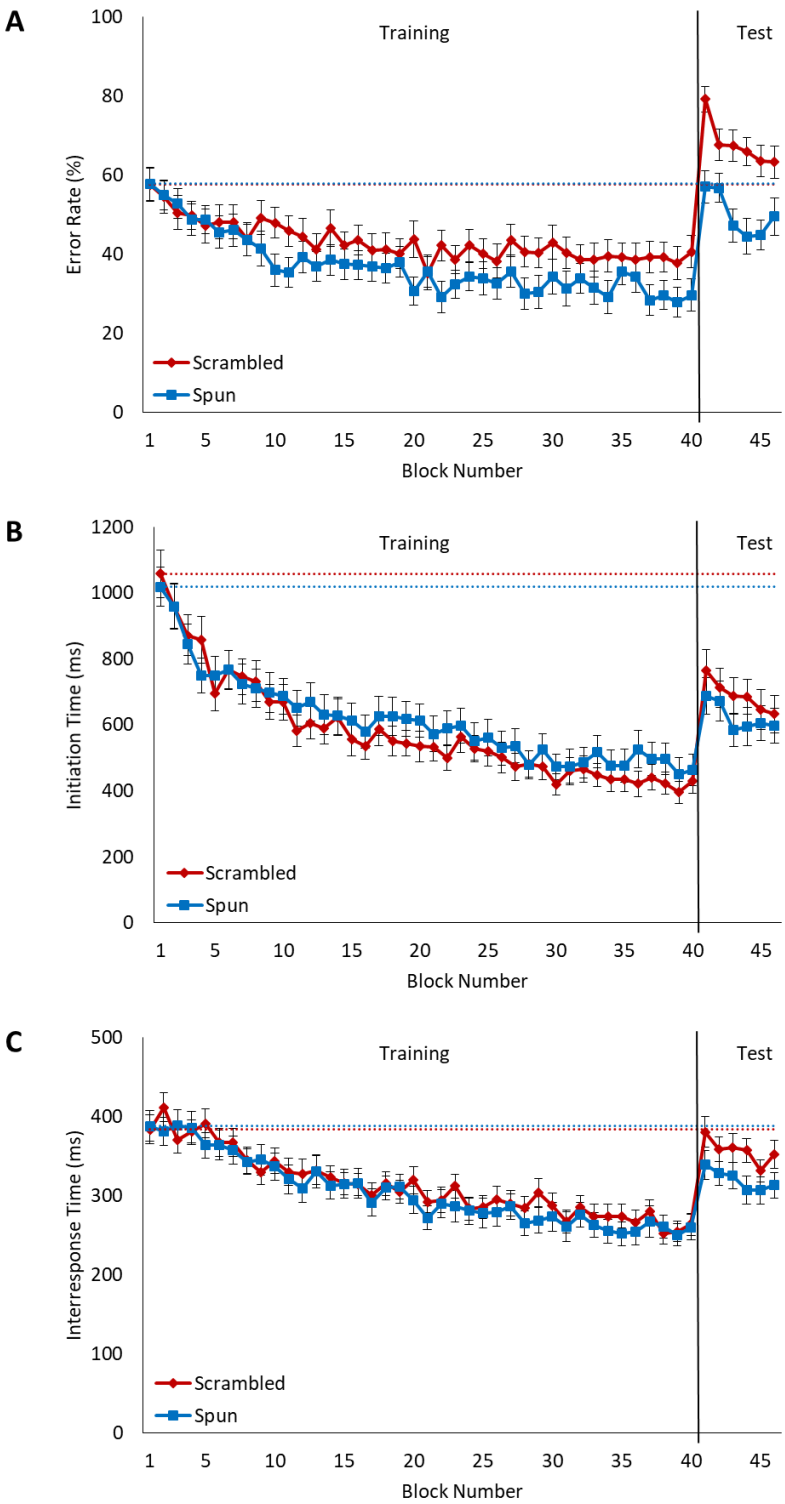

Figure 3. Experiment 2: Mean error rate (Panel A), IT (Panel B), and IRT (Panel C) for each list type in each of the 46 (40 training + 6 test) experiment blocks. The bars are standard errors of the means. The straight dotted lines are first block performance, and the color of these lines indicates the list type to which they belong.

Table 5

Experiment 2: ANOVA and simple main effect analyses for training and transfer effects.

DV	F	dfs	MSE	p	ε	η_p^2
			Number of Trained Lists Per Set			
Error Rate	0.796	1, 46	3.537	0.377	1.000	0.017
IT	1.080	1, 46	6,799,164	0.304	1.000	0.023
IRT	1.096	1, 46	679,452	0.301	1.000	0.023
			Spun vs. Scrambled Performance (Training)			
Error Rate	4.583	1, 46	0.861	0.038	1.000	0.091
IT	1.219	1, 46	914,766	0.275	1.000	0.026
IRT	0.545	1, 46	133,354	0.464	1.000	0.012
			Learning in Training (Spun)			
Error Rate	10.187	39, 1794	0.026	< 0.001	0.617	0.181
IT	27.628	39, 1794	30,039	< 0.001	0.405	0.375
IRT	23.021	39, 1794	3,763	< 0.001	0.632	0.334
			Learning in Training (Scrambled)			
Error Rate	4.208	39, 1794	0.026	< 0.001	0.712	0.084
IT	39.060	39, 1794	30,039	< 0.001	0.253	0.459
IRT	20.633	39, 1794	3,763	< 0.001	0.675	0.310
			Spun vs. Scrambled Performance (Test)			
Error Rate	36.386	1, 46	0.127	< 0.001	1.000	0.442
IT	4.358	1, 46	136,467	0.042	1.000	0.087
IRT	9.457	1, 46	20,750	0.004	1.000	0.171
			Presence of Transfer (Spun)			
Error Rate	0.340	1, 46	0.092	0.563	1.000	0.007
IT	76.449	1, 46	95,436	< 0.001	1.000	0.624
IRT	36.044	1, 46	13,788	< 0.001	1.000	0.439
			Presence of Transfer (Scrambled)			
Error Rate	42.882	1, 46	0.092	< 0.001	1.000	0.481
IT	49.295	1, 46	95,436	< 0.001	1.000	0.517
IRT	7.783	1, 46	13,788	0.008	1.000	0.145
			Completeness of Transfer (Spun)			
Error Rate	59.644	1, 46	0.087	< 0.001	1.000	0.559
IT	32.894	1, 46	83,976	< 0.001	1.000	0.417
IRT	46.220	1, 46	12,151	< 0.001	1.000	0.501
			Completeness of Transfer (Scrambled)			
Error Rate	135.838	1, 46	0.087	< 0.001	1.000	0.743
IT	119.555	1, 46	83,976	< 0.001	1.000	0.722
IRT	99.595	1, 46	12,151	< 0.001	1.000	0.684

Presence and completeness of transfer analyses are simple main effects analyses comparing test performance to initial and final training performance and use the MSE of the List Type X Start Portion and List Type X End Portion interactions, respectively.

A 2 (group) x 2 (list type) x 40 (training block) ANOVA was conducted on each performance measure to test for training effects. The main effect of list type and simple main effects of block number are shown in Table 5. Error rate was lower for spun lists during training (spun: 37.1%; scrambled: 43.1%), supported by a main effect of list type (Table 5), despite no pre-experimental difference in performance between the two list types (spun: 57.8%; scrambled: 57.6%; Table 6). There was no list type difference in IT (spun: 611.4 ms; scrambled: 577.3 ms) or IRT (spun: 303.6 ms; scrambled: 312.3 ms; Table 5). These findings replicate what was observed in Experiment 1: Consistent relative order improves the accuracy of recalling spun lists. The post-experiment survey revealed that 10 (20.8%) of the 48 participants noticed the consistent relative order in spun lists, and 13 (27.1%) noticed additional structure but not consistent relative order. I re-ran the error rate ANOVA, including spin detection and structure detection as additional factors. Structure detection did not affect the spun list advantage, $F(1, 44) = 0.009$, MSE = 0.140, $p = .926$, indicating again that the conscious detection of additional structure in spun lists is not necessary for participants to do better in these lists. Spin detection, on the other hand, magnified the spun list advantage, $F(1, 44) = 7.155$, MSE = 0.667, $p = .010$, suggesting that if participants happen to notice the consistent relative order in spun lists, they can exploit it to further improve their recall accuracy.

Transfer of training into the test portion was diagnosed with a combination of tests. The most critical test was on the difference in test portion performance in spun and scrambled lists – a main effect analysis on list type in the test portion. Although the spun lists in the test portion were new, the relative order of letters in these lists was the same as the relative order learned during training. If the item-to-item associations learned in the training portion transfer to test, then performance should be better on new spun lists than new scrambled lists. I also present the results of a simple main effect analysis for each list type that compared performance on the 6 test blocks to performance on the first 6 training blocks. These simple main effect analyses tested for the presence of transfer in each list type and inform why (or why not) test performance differed for the two list types.

Table 6

Experiment 2: T-tests for first block performance and power function parameters.

DV	t	df	M_{A-B}	SE_{A-B}	p	d	BF
Spun vs. Scrambled First Block Performance							
Error Rate	0.050	47	0.003	0.052	0.960	0.009	6.369 (N)
IT	-0.746	47	-39.456	52.897	0.459	-0.086	4.903 (N)
IRT	0.212	47	3.923	18.462	0.833	0.030	6.242 (N)
Spun vs. Scrambled Learning Rate (Training)							
Error Rate	1.970	47	0.076	0.039	0.027	0.360	1.083 (N)
IT	-1.065	47	-0.065	0.061	0.146	-0.208	3.744 (N)
IRT	0.279	47	0.020	0.070	0.391	0.058	3.815 (N)
Spun vs. Scrambled Starting Point (Test)							
Error Rate	-4.275	47	-0.163	0.038	< 0.001	-0.709	248.516 (A)
IT	-2.304	47	-86.434	37.516	0.013	-0.173	1.716 (A)
IRT	-2.011	47	-35.204	17.509	0.025	-0.254	1.008 (N)
Spun vs. Scrambled Learning Rate (Test)							
Error Rate	1.680	47	0.069	0.041	0.050	0.280	1.733 (N)
IT	0.400	47	0.033	0.082	0.345	0.082	5.911 (N)
IRT	1.529	47	0.128	0.084	0.067	0.331	2.156 (N)

Starting point analyses are conducted on the b parameters obtained from power function fits.

The results of these analyses are shown in Table 5. New spun lists were recalled more accurately and more quickly than new scrambled lists (error rate: 50.0% vs. 67.8%; IT: 632.8 ms vs. 688.0 ms; IRT: 319.9 ms vs. 356.8 ms), supported by significant main effects of list type in the test portion (Table 5). This finding is additional evidence that item-to-item associations are used in memory for serial order – it indicates that item-to-item associations were learned during training and transferred to the test portion.

The simple main effect analyses revealed that there was significant positive transfer in IT and IRT for both list types (spun test vs. spun training IT: 632.8 ms vs. 848.8 ms; scrambled test vs. scrambled training IT: 688.0 ms vs. 868.8 ms; spun test vs. spun training IRT: 319.9 ms vs. 378.7 ms; scrambled test vs. scrambled training IRT: 356.8 ms vs. 384.1 ms; Table 5). The list type differences in test portion IT and IRT indicate that the positive transfer was greater for spun lists. The presence of positive transfer in scrambled lists can be attributed to task-specific learning or to participants learning the letter sets. It emphasizes the importance of looking at the difference in performance between the list types rather than the list-type transfer effects alone.

On the other hand, there was negative transfer in error rate for scrambled lists (test: 67.8%; training: 51.3%), and no transfer in error rate for spun lists (test: 50.0%; training: 51.4%; Table 5). The list type difference in error rate seems to be the result of item-to-item associations counteracting interference. For both types of list, the position-to-item associations formed during training could interfere with the learning of the new lists, which presented novel position-to-item pairings. The scrambled lists suffered a reduction in accuracy in the test portion from this interference. The consistent relative order in spun lists seemed to protect spun list performance from this interference.

Power function fits. Four functions were fit to each participant's data – one for each list type in each of the training and test portions. The training and test functions were fit separately. The spun training and scrambled training functions were fit simultaneously, and the spun test and scrambled test functions were fit simultaneously.

If item-to-item associations transfer from training, then the starting point – controlled by *b* parameter because *a* is fixed – should be lower for the spun test function than the scrambled test function. Because the transfer of item-to-item associations predicts a difference in starting point, I allowed the *b* parameter to differ between the spun and scrambled test functions. I allowed the *c* parameter to differ between the two test functions as well, but a difference in the *c* parameter is less diagnostic of the transfer of item-to-item associations.[2] Four parameters were estimated for each participant's test functions: a *b* parameter for spun lists, a *b* parameter for scrambled lists, a *c* parameter for spun lists, and a *c* parameter for scrambled lists. Like Experiment 1, three parameters were estimated for each participant's training functions.

For the training functions, average R^2 was .301, .407, and .402 for error rate, IT, and IRT respectively. The average shared *b* parameters were 65.1, 324.4, and 324.2, respectively. In the training portion, learning rate was significantly faster for spun lists in error rate (spun: .278; scrambled: .202). Learning rate was numerically but not significantly faster in IT (spun: .433;

[2] The *c* parameter for both functions would be close to zero if very little learning occurred in the test portion. If transfer of training was perfect, a pattern of data might emerge where neither type of list improved during the test portion, yet spun list performance was better than scrambled list performance. The t-test on the *c* parameter would misleadingly indicate that there is no difference in transfer. The t-test on the *b* parameter, on the other hand, would indicate a difference in transfer.

scrambled: .357) and IRT (spun: .427; scrambled: .407; Table 6). Like the ANOVA, the test on learning rate indicates an advantage for spun lists during training.

For the test functions, average R^2 was .529, .399, and .417, respectively. In the test portion, the starting point of performance was lower for spun lists in error rate (spun: 62.3; scrambled: 78.6), IT (spun: 632.6; scrambled: 719.1), and IRT (spun: 218.1; scrambled: 253.3), supported by a significant difference in the *b* parameter between list types (Table 6). The test portion learning rate was also higher for spun lists in error rate (spun: .247; scrambled: .178; Table 6). Like the ANOVA, these tests indicate better performance on – and thus better transfer to – spun lists in the test portion.

Chapter Summary

In this chapter, I presented two experiments that pit spun lists against scrambled lists. These experiments more sharply tested for item-to-item associations than past research (e.g., Kahana, Mollison, & Addis, 2010). Any additional learning in spun lists could be attributed to the presence of consistent relative order, which allowed for the development of item-to-item associations. Spun lists were learned more quickly, and training on spun lists transferred to new lists that shared the same relative order. These findings suggest that participants formed item-to-item associations and used them to retrieve letters in the spun lists. The way spun list performance diverged from scrambled list performance suggested that the learning of item-to-item associations followed a closed-loop learning rule.

CHAPTER 3

THE LOCUS OF ITEM-TO-ITEM ASSOCIATIONS

Chapter 1 demonstrated that people do use item-to-item associations in serial recall – specifically when relative position is consistent and serial position is inconsistent. However, tasks are not pure measures of cognitive processes, and the purpose of this paper is to determine whether item-to-item associations are used in serial memory. Serial recall tasks tap into the memory system, but they tap into other systems as well. I did not relay the lists to the participants telepathically; the lists in my experiments were presented visually, so they had to pass through the visual system before reaching the memory system. Likewise, the participants did not relay their responses to me telepathically; they had to type their responses, so information had to be passed through the motor system.

From stimulus to response, the lists in my experiments are represented in multiple different forms: first as iconic representations in perception, then as categorical representations in memory, and finally as motor commands in the motor system. These representations are coupled – each letter has one associated visual representation and one associated response – so any consistencies in order (or lack thereof) apply to all of the representations. The item-to-item associations may have been formed between items in memory, or they may have been formed between iconic representations in perception or motor commands in the motor system. As a result, my ability to conclude that these are item-to-item associations in serial memory has so far been occluded.

In this chapter I address where in the stream of processing the item-to-item associations that produce the spun list advantage are located. My strategy is to decouple the perceptual, memory, and motor representations of the letters by changing one of the letter representations – either the perceptual representations or the motor representations. In the experiments in this chapter, participants trained on spun lists, I changed the perceptual or motor representations of the letters, and then participants tested on the same lists of letters to see if the advantage for spun lists persists. If changing the representations causes the advantage to go away, then the associations were formed between whatever representations were changed. If the advantage does not go away, the associations were not formed between the changed representations.

Experiment 3

The methods of Experiment 1 and Experiment 2 were adapted from Lindsey and Logan (2019), the primary difference being the exposure duration of the lists. In my experiments, the lists were displayed for 500 ms. In Lindsey and Logan's (2019) experiments, the lists remained on screen until the participant finished typing them. In both studies, the participants eventually had to type the sequence of letters in order, and learning was greater for the spun sequences. The learning advantage observed in both studies may reflect the formation of associations between motor commands. In both studies, keystroke transitions were consistent for spun sequences but not scrambled sequences.

In Experiment 3, I attempted to remove any typing-specific contributions to the learning observed for spun lists by decoupling memory representations and motor representations. Participants first trained on a set of spun and scrambled lists using one response modality – either typing the responses or speaking the responses into a microphone – and then they were tested on the same lists using a different response modality (e.g., speaking the lists if they typed the lists during training). I expected to see a learning advantage for spun lists in the training portion. However, if the item-to-item associations that produced this advantage were isolated to the motor system, then this advantage should disappear in the test portion because the response method changed (from manual to oral, or vise-versa). Participants who typed lists during training should be no better at speaking those same lists during the test portion. On the other hand, if item-to-item associations were not isolated to the motor system, then participants who typed lists during training should also be better at speaking those lists during the test portion.

Method

Participants. 24 participants were recruited for this experiment using the same selection criteria as previous experiments. Participants were scheduled in 1.5 hour timeslots and received $18 or course credit as compensation.

Apparatus and stimuli. Participants practiced the same two sets of lists – one spun and one scrambled – using both typed and spoken responses. The pool of selectable letters in this

experiment was reduced to exert more control over the phonological characteristics of the lists. I omitted all vowel letters, including ‘y,’ to prevent the formation of nonsense syllables. I omitted ‘w’ because it has two syllables when pronounced alone. I omitted all letters with the “-ee” sound, excluding ‘z’ because some pronounce it “zed,” to lessen the likelihood of phonological confusion errors. The remaining pool of 12 letters available for selection was ‘f’, ‘h’, ‘j’, ‘k’, ‘l’, ‘m’, ‘n’, ‘q’, ‘r’, ‘s’, ‘x’, and ‘z.’ For each participant, six of these letters were randomly selected to be the letters of the spun set, and the remaining 6 letters become the scrambled set. Six spun lists and 6 scrambled lists were made from these 2 sets in the same manner as previous experiments. Example lists are shown in Table 7.

Table 7
Construction of letter sequences in Experiment 3.

Spun Set							Scrambled Set						
Item 1	q	r	s	f	h	j	Item 7	k	l	x	n	m	z
Item 2	r	s	f	h	j	q	Item 8	l	n	k	z	x	m
Item 3	s	f	h	j	q	r	Item 9	n	z	l	m	k	x
Item 4	f	h	j	q	r	s	Item 10	z	m	n	x	l	k
Item 5	h	j	q	r	s	f	Item 11	m	x	z	k	n	l
Item 6	j	q	r	s	f	h	Item 12	x	k	m	l	z	n

These letters serve as examples. Each subject received a different random selection of letters. When presented on the computer screen, the spaces between letters were removed. Each of the items was presented during training and test.

Procedure. Like Experiment 2, there was a training portion and a test portion to this experiment. Half of the participants (12) practiced typing each of the 12 lists 40 times in the training portion and then practiced speaking each of the lists 10 times in the test portion. There was a short break between the training and test portion, during which these participants were instructed to speak the letters they remember into a microphone instead of typing them when the “Response:” screen appeared. Like before, they pressed the spacebar to finish the trial and move on to the next trial. The other half of participants practiced speaking the 12 lists during the training portion and practiced typing the lists during the test portion. All participants completed 600 trials (480 training and 120 test), and self-paced breaks were provided every 120 trials.

Results and Discussion

Spoken trials were scored manually. I listened to each of the sound files, transcribed the letter responses, and then scored the accuracy of the transcribed responses by comparing them to the letters presented on each trial. Due to difficulties in detecting the start and end times of each utterance in the sound files, IT and IRT were not distinguished in this experiment. These timing measures were replaced by a measure of the total time taken on each trial, measured from the onset of the response screen to the spacebar press. The time taken on a trial was only included in averages if it was less than 10 sec.

Mean error rate and time taken were calculated for each participant (24) in each group (2) on each list type (2) in each block of the experiment (50). The grouping variable – the order in which participants did typing and speech – did not have a significant effect on performance (Table 8), so the reported analyses average over group. Training portion means were analyzed using and 2 (group) x 2 (list type) x 40 (training block) ANOVA, and the list type main effects and block number simple main effects are shown in Table 8. Test portion means were analyzed using a 2 (group) x 2 (list type) x 10 (test block) ANOVA (for the comparison of spun and scrambled test lists) and 2 (group) x 2 (portion) x 10 (test block) ANOVA for each list type (for testing transfer in each list type). The main effects of list type, simple main effects of the presence of transfer (first 10 training blocks vs. the 10 test blocks), and simple main effects of the completeness of transfer (last 10 training blocks vs. the 10 test blocks) are also shown in Table 8. Mean performance for each of the experiment blocks, averaging over group, is shown in Figure 4.

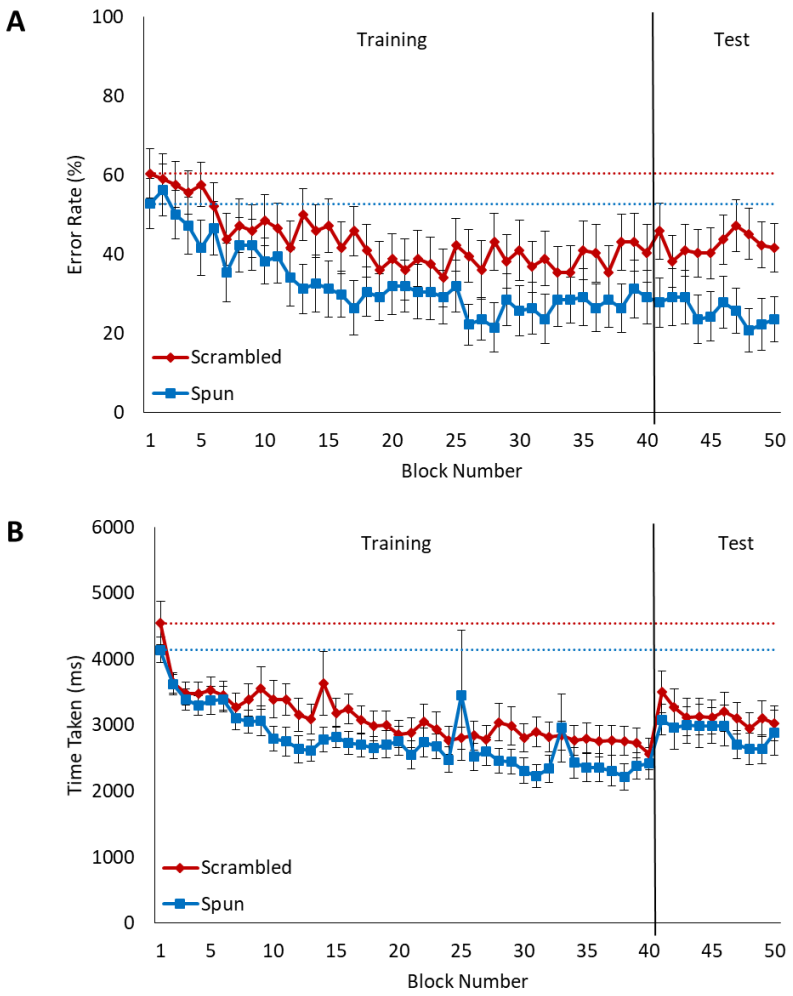

Figure 4. Experiment 3: Mean error rate (Panel A) and time taken (Panel B) for each list type in each of the 50 (40 training + 10 test) experiment blocks. The bars are standard errors of the means. The straight dotted lines are first block performance, and the color of these lines indicates the list type to which they belong.

Although error rate was numerically lower for spun lists than scrambled lists (33.1% vs. 43.5%), this difference did not reach significance (Table 8). This is likely an issue with power: Reducing the number of letter sets to 2 cut the number of trials per block in half relative to Experiment 1. Unlike previous experiments, a time-related advantage for spun lists was observed: The time taken on spun lists was significantly lower than on scrambled lists (2,764.9 ms vs. 3,099.2 ms; Table 8).

Table 8

Experiment 3: ANOVA and simple main effect analyses for training and transfer effects.

DV	F	dfs	MSE	p	ε	η_p^2
			Response Method Order			
Error Rate	2.500	1, 22	4.911	0.128	1.000	0.102
Time Taken	1.708	1, 22	51,276,068	0.205	1.000	0.072
			Spun vs. Scrambled Performance (Training)			
Error Rate	2.989	1, 22	1.737	0.098	1.000	0.120
Time Taken	4.686	1, 22	11,446,454	0.042	1.000	0.176
			Learning in Training (Spun)			
Error Rate	7.196	39, 858	0.024	< 0.001	0.583	0.246
Time Taken	6.707	39, 858	662,992	< 0.001	0.050	0.234
			Learning in Training (Scrambled)			
Error Rate	5.082	39, 858	0.024	< 0.001	0.567	0.188
Time Taken	5.141	39, 858	662,992	< 0.001	0.211	0.189
			Spun vs. Scrambled Performance (Test)			
Error Rate	7.074	1, 22	0.499	0.014	1.000	0.243
Time Taken	1.341	1, 22	6,295,268	0.259	1.000	0.057
			Presence of Transfer (Spun)			
Error Rate	65.544	1, 22	0.072	< 0.001	1.000	0.749
Time Taken	10.895	1, 22	2,068,954	0.003	1.000	0.331
			Presence of Transfer (Scrambled)			
Error Rate	124.392	1, 22	0.072	< 0.001	1.000	0.850
Time Taken	27.075	1, 22	2,068,954	< 0.001	1.000	0.552
			Completeness of Transfer (Spun)			
Error Rate	1.955	1, 22	0.034	0.176	1.000	0.082
Time Taken	16.582	1, 22	1,729,417	< 0.001	1.000	0.430
			Completeness of Transfer (Scrambled)			
Error Rate	64.320	1, 22	0.034	< 0.001	1.000	0.745
Time Taken	0.955	1, 22	1,729,417	0.339	1.000	0.042

Table 9

Experiment 3: T-tests for first block performance and power function parameters.

DV	t	df	M_{A-B}	SE_{A-B}	p	d	BF
Spun vs. Scrambled First Block Performance							
Error Rate	-0.891	23	-0.076	0.086	0.382	-0.248	3.258 (N)
Time Taken	-1.239	23	-402.556	324.83	0.228	-0.295	2.357 (N)
Spun vs. Scrambled Learning Rate (Training)							
Error Rate	2.347	23	0.158	0.067	0.014	0.671	2.074 (A)
Time Taken	1.765	23	0.158	0.089	0.045	0.411	1.219 (N)
Spun vs. Scrambled Starting Point (Test)							
Error Rate	-1.894	23	-0.135	0.071	0.035	-0.458	1.009 (N)
Time Taken	-1.361	23	-638.161	468.756	0.093	-0.317	2.058 (N)
Spun vs. Scrambled Learning Rate (Test)							
Error Rate	2.711	23	0.203	0.075	0.006	0.783	4.027 (A)
Time Taken	0.111	23	0.016	0.144	0.456	0.035	4.632 (N)

Spun lists were recalled more accurately during the test portion (spun: 25.4% error rate; scrambled: 42.6% error rate). Transfer of training was perfect to the new response modality; test performance was not significantly different than performance at the end of training (test: 25.4%; training end: 27.7%; Table 8). The advantage for spun lists is consistent with people using item-to-item associations. Moreover, the persistence of this advantage after switching to a different response method in the test portion suggests that these associations are not tied to the motor system.

On the other hand, the time advantage for spun lists disappeared after switching response method (spun: 2,888.9 ms; scrambled: 3,154.1 ms; Table 8), suggesting that the associations are response method specific. However, the time advantage is an anomaly in the context of the previous two experiments (and experiments after this one). If the time advantage is real, then it may be that the associations that produce the accuracy-related spin advantage (seen in the current experiments) exist between representations in memory, while the associations that produce the time-related spin advantage (seen in this experiment and in Lindsey & Logan, 2019) exist between representations in the motor system.

The post-experiment interviews revealed that 2 participants (8.3%) noticed consistent relative order in spun lists, and another 10 participants (41.7%) noticed that there was additional structure in the spun lists. However, the error rate advantage for spun lists was not affected by

the detection of spin, $F(1, 20) = 0.213$, MSE = 2.347, $p = .649$, or the detection of structure, $F(1, 20) = 0.923$, MSE = 2.180, $p = .348$.

Power function fits. Average R^2 from fits to the training data were .416 and .492 for error rate and time taken, respectively. The average shared *b* parameter was 69.3 for error rate and 5,031.2 for time taken. In the training portion, learning rate was faster for spun lists in both error rate (spun: .388; scrambled: .231; there were no power issues here) and time taken (spun: .498; scrambled: .340; Table 9).

The average R^2 for the test data were .485 and .415 for error rate and time taken, respectively. In the test portion, the starting point for error rate was lower for spun lists (spun: 32.6; scrambled: 46.1), and the learning rate for error rate was higher for spun lists (spun: .261; scrambled: .058; Table 9). In the time taken to recall spun lists, starting point was numerically lower (spun *b*: 3,754.3; scrambled *b*: 4,392.5), and learning rate was numerically higher (spun *c*: .454; scrambled *c*: .438). However, these differences were not significant (Table 9). The tests on power function fits were generally consistent with the ANOVA results. The test on the training portion learning rate seemed to a more powerful test than the ANOVA, however.

Experiment 4

The lists in current experiments have been presented visually, but people tend to recode visually presented letters when covert articulation is not prevented (Scarborough, 1972). As a result, a list is likely represented first as a sequence of visual letter shapes, then as a sequence of categorical letter identities. The learning advantage observed for spun lists may reflect the strengthening of associations among visual representations in perception – allowing people to more quickly or more accurately encode the letters on the screen – or the strengthening of associations among phonological representations in working memory – allowing letters in memory to be remembered more quickly or more accurately.

In Experiment 4, I tested whether the item-to-item associations formed in spun lists are isolated to perception by manipulating the visual representations of the lists. First, participants practiced on spun and scrambled lists with one letter case (uppercase or lowercase), and then they tested on the same lists with the opposite letter case. The uppercase and lowercase representations of a letter are sometimes visually distinct, but they are united by a common letter

identity. In the training portion, participants should again learn the spun lists faster over practice. If item-to-item associations are isolated to perception, then changing the visual representations of the lists should disrupt the learning from the training portion. The advantage for spun lists should disappear in the test portion. However, if the associations are not isolated to perception, then changing the visual representation should not disrupt learning, and the spun list advantage should persist.

Method

Participants. 24 participants were recruited for this experiment, using the same selection criteria as previous experiments. Participants were scheduled in 1 hour timeslots and received $12 or course credit as compensation.

Apparatus and stimuli. Participants practiced the same two sets of spun and scrambled lists in the training and test portions. I trimmed the pool of selectable letters to those with more distinct lowercase and uppercase forms. I omitted letters with very similar lowercase and uppercase forms (e.g., ‘s’ and ‘S’) and letters with lowercase forms that look like other letters (e.g., ‘l’ and ‘I’). I also omitted ‘a’ and ‘e’ like the previous experiments to lessen the likelihood that participants form nonsense syllables. The remaining pool of 12 letters was ‘b’, ‘d’, ‘g’, ‘h’, ‘j’, ‘m’, ‘n’, ‘q’, ‘r’, ‘t’, ‘u’, and ‘y.’ Six of these letters were randomly selected to be the letters of the spun set, and the remaining 6 letters become the scrambled set. Six spun lists and 6 scrambled lists were made from these 2 sets. Example lists are shown in Table 10.

Procedure. Participants were explicitly instructed to type all letters in lowercase, because the Shift and Caps Lock keys were disabled. Like Experiment 3, participants practiced the same 12 lists 50 times over the course of the experiment. For half of the participants (12) the first 40 exposures were in lowercase, and the last 10 exposures were in uppercase. For the other half, the first 40 exposures were in uppercase, and the last 10 exposures were in lowercase. Like Experiment 1 and Experiment 2, participants always typed their responses. They completed 600 (480 training and 120 test) trials and had the opportunity to take self-paced breaks every 120 trials.

Table 10

Construction of letter sequences in Experiment 4.

Spun Set								Scrambled Set						
							Group 1							
Training								Training						
	Item 1	q	g	b	n	r	t		Item 7	d	u	j	h	m	y
	Item 2	g	b	n	r	t	q		Item 8	u	h	d	y	j	m
	Item 3	b	n	r	t	q	g		Item 9	h	y	u	m	d	j
	Item 4	n	r	t	q	g	b		Item 10	y	m	h	j	u	d
	Item 5	r	t	q	g	b	n		Item 11	m	j	y	d	h	u
	Item 6	t	q	g	b	n	r		Item 12	j	d	m	u	y	h
Test								Test						
	Item 1	Q	G	B	N	R	T		Item 7	D	U	J	H	M	Y
	Item 2	G	B	N	R	T	Q		Item 8	U	H	D	Y	J	M
	Item 3	B	N	R	T	Q	G		Item 9	H	Y	U	M	D	J
	Item 4	N	R	T	Q	G	B		Item 10	Y	M	H	J	U	D
	Item 5	R	T	Q	G	B	N		Item 11	M	J	Y	D	H	U
	Item 6	T	Q	G	B	N	R		Item 12	J	D	M	U	Y	H
							Group 2							
Training								Training						
	Item 1	Q	G	B	N	R	T		Item 7	D	U	J	H	M	Y
	Item 2	G	B	N	R	T	Q		Item 8	U	H	D	Y	J	M
	Item 3	B	N	R	T	Q	G		Item 9	H	Y	U	M	D	J
	Item 4	N	R	T	Q	G	B		Item 10	Y	M	H	J	U	D
	Item 5	R	T	Q	G	B	N		Item 11	M	J	Y	D	H	U
	Item 6	T	Q	G	B	N	R		Item 12	J	D	M	U	Y	H
Test								Test						
	Item 1	q	g	b	n	r	t		Item 7	d	u	j	h	m	y
	Item 2	g	b	n	r	t	q		Item 8	u	h	d	y	j	m
	Item 3	b	n	r	t	q	g		Item 9	h	y	u	m	d	j
	Item 4	n	r	t	q	g	b		Item 10	y	m	h	j	u	d
	Item 5	r	t	q	g	b	n		Item 11	m	j	y	d	h	u
	Item 6	t	q	g	b	n	r		Item 12	j	d	m	u	y	h

These letters serve as examples. Each subject received a different random selection of letters. When presented on the computer screen, the spaces between letters were removed.

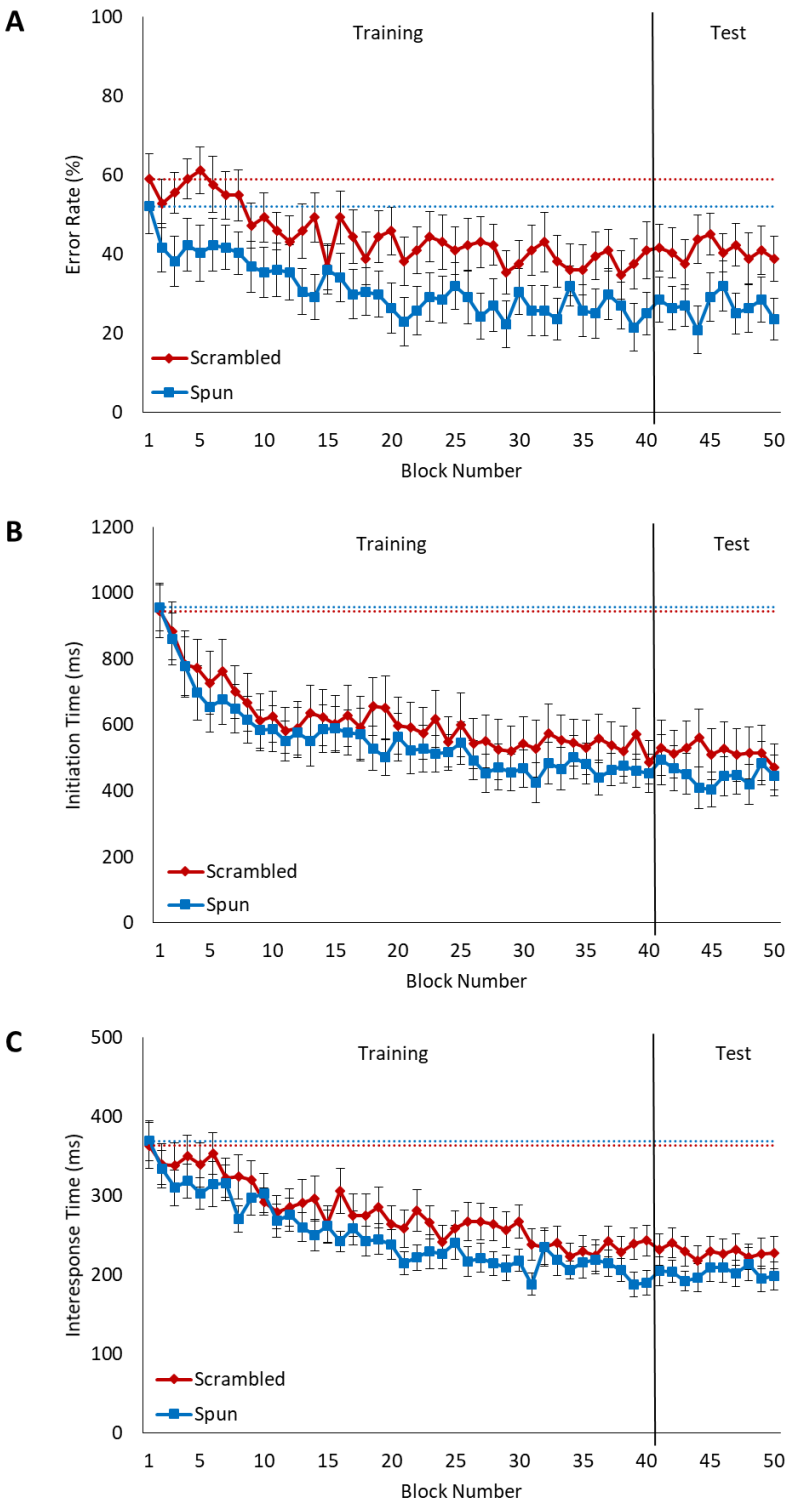

Figure 5. Experiment 4: Mean error rate (Panel A), IT (Panel B), and IRT (Panel C) for each list type in each of the 50 (40 training + 10 test) experiment blocks. The bars are standard errors of the means. The straight dotted lines are first block performance, and the color of these lines indicates the list type to which they belong.

Table 11

Experiment 4: ANOVA and simple main effect analyses for training and transfer effects.

DV	F	dfs	MSE	p	ε	η_p^2
			Number of Trained Lists Per Set			
Error Rate	0.266	1, 22	5.947	0.611	1.000	0.012
IT	0.107	1, 22	9,721,479	0.747	1.000	0.005
IRT	0.028	1, 22	660,432	0.869	1.000	0.001
			Spun vs. Scrambled Performance (Training)			
Error Rate	11.358	1, 22	0.742	0.003	1.000	0.340
IT	1.727	1, 22	985,374	0.202	1.000	0.073
IRT	3.676	1, 22	110,347	0.068	1.000	0.143
			Learning in Training (Spun)			
Error Rate	4.500	39, 858	0.026	< 0.001	0.443	0.170
IT	15.058	39, 858	20,424	< 0.001	0.427	0.406
IRT	16.878	39, 858	2,798	< 0.001	0.475	0.434
			Learning in Training (Scrambled)			
Error Rate	4.930	39, 858	0.026	< 0.001	0.774	0.183
IT	11.626	39, 858	20,424	< 0.001	0.329	0.346
IRT	13.148	39, 858	2,798	< 0.001	0.509	0.374
			Spun vs. Scrambled Performance (Test)			
Error Rate	18.978	1, 22	0.128	< 0.001	1.000	0.463
IT	2.340	1, 22	261,965	0.140	1.000	0.096
IRT	3.487	1, 22	23,630	0.075	1.000	0.137
			Presence of Transfer (Spun)			
Error Rate	23.729	1, 22	0.105	< 0.001	1.000	0.519
IT	127.814	1, 22	63,301	< 0.001	1.000	0.853
IRT	181.003	1, 22	8,197	< 0.001	1.000	0.892
			Presence of Transfer (Scrambled)			
Error Rate	92.637	1, 22	0.105	< 0.001	1.000	0.808
IT	172.246	1, 22	63,301	< 0.001	1.000	0.887
IRT	254.100	1, 22	8,197	< 0.001	1.000	0.920
			Completeness of Transfer (Spun)			
Error Rate	0.143	1, 22	0.033	0.709	1.000	0.001
IT	2.729	1, 22	15,628	0.113	1.000	0.110
IRT	1.426	1, 22	2,715	0.245	1.000	0.061
			Completeness of Transfer (Scrambled)			
Error Rate	53.517	1, 22	0.033	< 0.001	1.000	0.299
IT	67.908	1, 22	15,628	< 0.001	1.000	0.755
IRT	45.265	1, 22	2,715	< 0.001	1.000	0.673

Results and Discussion

Data analysis in this experiment was identical to Experiment 3. Participant group had no effect on any of the dependent measures, so the reported analyses average over group. Mean performance averaging over participants and groups is shown in Figure 5. There was an error rate advantage for spun lists during training (spun: 31.5%; scrambled: 44.8%), but no advantage for IT (spun: 557.1 ms; scrambled: 616.7 ms) or IRT (spun: 249.3 ms; scrambled: 278.4 ms; Table 11). There was also an error rate advantage for spun lists during the test portion (spun: 26.7%; scrambled: 41.0%), and transfer was perfect to the new spun lists (test: 26.7%; training end: 26.1%; Table 11). Switching letter case did not abolish the spun list advantage, suggesting that the item-to-item associations used to remember these lists are not isolated to perception.

Table 12
Experiment 4: T-tests for first block performance and power function parameters.

DV	t	df	M_{A-B}	SE_{A-B}	p	d	BF
Spun vs. Scrambled First Block Performance							
Error Rate	-1.082	23	-0.069	0.064	0.290	-0.213	2.760 (N)
IT	0.258	23	11.882	46.123	0.799	0.032	4.519 (N)
IRT	0.223	23	5.783	25.915	0.825	0.043	4.554 (N)
Spun vs. Scrambled Learning Rate (Training)							
Error Rate	3.336	23	0.231	0.069	0.001	0.897	14.020 (A)
IT	0.794	23	0.080	0.101	0.218	0.227	3.503 (N)
IRT	1.308	23	0.120	0.092	0.102	0.376	2.186 (N)
Spun vs. Scrambled Starting Point (Test)							
Error Rate	-4.223	23	-0.158	0.037	< 0.001	-0.675	95.042 (A)
IT	0.375	23	28.425	75.867	0.356	0.058	4.369 (N)
IRT	-1.521	23	-23.843	15.671	0.071	-0.286	1.695 (N)
Spun vs. Scrambled Learning Rate (Test)							
Error Rate	0.510	23	0.030	0.058	0.307	0.166	4.138 (N)
IT	1.207	23	0.150	0.124	0.120	0.375	2.438 (N)
IRT	0.367	23	0.044	0.121	0.359	0.103	4.381 (N)

The post-experiment interviews revealed that 2 participants (8.3%) noticed consistent relative order in spun lists, and another 10 participants (41.7%) noticed that there was additional structure in the spun lists. The error rate advantage for spun lists was not affected by the

detection of structure, $F(1, 20) = 0.029$, MSE = 0.841, $p = .867$, but the detection of spin magnified the error rate advantage, $F(1, 20) = 6.190$, MSE = 0.657, $p = .022$. Noticing additional structure in spun lists is not a prerequisite for obtaining a spun list advantage, but noticing consistent relative order can modulate the effect.

Power function fits. Average R^2 from fits to the training data were .318, .413, and .437 for error rate, IT, and IRT, respectively. The average shared *b* parameters were 62.6, 999.2, and 269.6, respectively. In the training portion, learning rate was significantly faster for spun lists in error rate (spun: .386; scrambled: .155). Learning rates were numerically but not significantly faster in IT (spun: .573; scrambled: .492) and IRT (spun: .517; scrambled: .397; Table 12).

The average R^2 for the test data were .283, .315, and .372, respectively. In the test portion, the starting point for error rate was lower for spun lists (spun: 25.8; scrambled: 41.6; Table 12). The learning rate for error rate did not differ between the list types, but this seems to be the result of very little learning occurring in the test portion (spun: .113; scrambled: .084; Table 12). Neither starting point nor learning rate were significantly different in IT (spun *b*: 449.5; scrambled *b*: 421.1; spun *c*: .404; scrambled *c*: .255; Table 12) or IRT (spun *b*: 71.3; scrambled *b*: 95.1; spun *c*: .392; scrambled *c*: .347; Table 12). The tests on power function fits were consistent with the ANOVA results – item-to-item associations supported more accurate retrieval in the training and test portions.

Chapter Summary

I presented two experiments in this chapter that tested whether the item-to-item associations used in serial recall are formed between memory representations. In the serial recall task, the perceptual, memory, and motor representations of the letters are normally coupled, and item-to-item associations could form between any of these representations. I decoupled these representations by changing how participants gave letter responses (Experiment 3) or how the letters appeared on the computer screen (Experiment 4) in the middle of the task. Decoupling the representations in these ways did not get rid of the spun list advantage, so the associations are not isolated to the motor system or to perception. The item-to-item associations used in serial recall are associations formed between representations of the letters in memory.

CHAPTER 4

DETECTION AND STRATEGIES

Kahana, Mollison, and Addis (2010) acknowledged that the learning observed in spun lists could arise from people detecting the structure of spun lists and developing a strategy that makes use of this structure. Detection of structure and strategy use have been concerns in this paper as well. Spun lists are inherently more structured than scrambled lists, so the advantage for spun lists might reflect participants using strategies that exploit this additional structure. The presented experiments have so far relied on post-hoc tests of self-reported detection to address this concern. These tests were not ideal – self-reports of cognitive processes are sometimes suspect (Nisbett & Wilson, 1977), the tests tended to have very lopsided sample sizes (few people could detect the consistent relative order in spun lists), and no attention was given to how the reported strategies themselves affected performance.

The goal of the current chapter is to address the effects of detection and strategies in a more rigorous manner. I present an experiment aimed at manipulating the ability of participants to detect structure in spun lists. Specifically, in this experiment I attempted to make detection more difficult, and I compared the results of this experiment to the results of Experiment 1 to determine how this added difficulty affected the spun list advantage. At the end of the chapter I analyze how different strategies affected performance and the spun list advantage. If the spun list advantage is a sharp measure of the use of item-to-item associations, then it should be relatively insensitive to the difficulty of detection and to the use of different memorization strategies.

Experiment 5

In Experiment 5, I aimed to make it more difficult to detect the additional structure in spun lists. I nested the spun items in the middle of the list, flanked by “anchor” items that did not change. Making detection more difficult should result in fewer people who detect the structure in spun lists. If the spun list advantage reflects detection of structure, then the spun list advantage should disappear.

Nesting the spun letters in the middle of the list does not affect detection alone. Keeping the end letters consistent and spinning the inner letters causes each of the end items to be

associated with each of the inner items. The nesting manipulation may cause interference in the item-to-item associations, and this interference may also cause the spun list advantage to disappear.

It is necessary to distinguish the predictions of detection and interference. Up to this point, I have focused on an aggregate measure of the spun list advantage, but it is possible to distinguish the predictions of detection and interference by considering what this advantage means at the level of individual participants. Not every participant in the presented experiments showed the spun list advantage – spun lists were more easily recalled than scrambled lists for a majority, but not all. Of the participants that did show the spun list advantage, the magnitude of this advantage varied from person to person. The aggregate spun list advantage is both a measure of the proportion of people who show the advantage – a higher proportion leads to a larger aggregate advantage – and a measure of the magnitude of the advantage in people who show the advantage – a higher magnitude leads to a higher aggregate advantage.

If detection produces the aggregate spun list advantage, it should do so through the proportion of people who show a spun list advantage. People who detect the structure in spun lists should show the advantage, and people who do not should not show the advantage. Only those who are using item-to-item associations for retrieval – those who show a spun list advantage – should be affected by interference in the item-to-item associations. Thus, interference affects the magnitude of the spun list advantage in those who show an advantage, not the proportion of people who show the advantage.

To preface the results, in Experiment 5 the aggregate spun list advantage disappeared, so I combined the data of Experiment 1 and Experiment 5 and treated experiment as a grouping variable to diagnose why the effect disappeared. This grouping variable reflects differences in the difficulty to detect structure in spun lists (easy in Experiment 1, hard in Experiment 5), and differences in interference in the item-to-item associations (none in Experiment 1, some in Experiment 5). If the likelihood of detection caused the disappearance, then fewer people should show a spun list advantage in Experiment 5 than Experiment 1. If interference caused the disappearance, then the magnitude of the spun list advantage should be lower for those who showed an advantage.

Method

Participants. I recruited 24 participants for this experiment, using the same selection criteria as previous experiments. Participants were compensated $12 or course credit for 1 hour of participation.

Apparatus and stimuli. 4 sets of 6 six letters were randomly selected like in Experiment 1. For each set, 2 letters were randomly selected to be anchor letters; 1 of them always appeared in position 1 of the list, and the other always appeared in position 6 of the list. The inner 4 letters were either spun or scrambled. This produced 4 spun lists or 4 scrambled lists per set, for a total of 8 spun lists and 8 scrambled lists. Example lists are shown in Table 13.

Table 13
Construction of letter sequences in Experiment 5.

Spun Set								Scrambled Set						
First Half								First Half						
	Item 1	t	p	g	u	b	k		Item 5	q	f	w	m	s	l
	Item 2	t	g	u	b	p	k		Item 6	q	w	s	f	m	l
	Item 3	t	u	b	p	g	k		Item 7	q	s	m	w	f	l
	Item 4	t	b	p	g	u	k		Item 8	q	m	f	s	w	l
Second Half								Second Half						
	Item 1	y	z	c	n	r	d		Item 5	x	h	o	i	j	v
	Item 2	y	c	n	r	z	d		Item 6	x	o	j	h	i	v
	Item 3	y	n	r	z	c	d		Item 7	x	j	i	o	h	v
	Item 4	y	r	z	c	n	d		Item 8	x	i	h	j	o	v

These letters serve as examples. Each subject received a different random selection of letters. When presented on the computer screen, the spaces between letters were removed.

Procedure. Aside from the number of trials and the spacing of breaks, the procedure of Experiment 5 was identical to Experiment 1. Participants practiced 8 lists in the first half and a different 8 lists in the second half – one set of spun lists and one set of scrambled lists per half. They practiced each list 40 times. In each half, participants completed 320 trials, for a total of 640 trials. Self-paced breaks were provided every 160 trials.

Results and Discussion

Mean performance over blocks in Experiment 5 is shown in Figure 6. There was a numerical advantage for spun lists in error rate, although it was slight and not significant (spun: 40.2%; scrambled: 45.1%; Table 14). Nesting the spun letters in the middle of the list caused the aggregate spun list advantage to disappear.

Only 2 of the 24 (8.3%) participants detected spin in this experiment, and only 2 more (8.3%) detected additional structure in spun lists. These numbers are down from 8 and 9, respectively, in Experiment 1. Logistic regression analyses, treating experiment as a predictor of spin detection and structure detection, indicated that these were significant differences, Wald = 3.965, df = 1, B = -1.705, S.E. = 0.856, p =.046 and Wald = 14.445, df = 1, B = -3.285, S.E. = 0.864, p = < .001, respectively. Nesting the spun letters in the middle of the list did have the intended effect of reducing the number of people that noticed additional structure in spun lists. In spite of this, the same number of people (16 of the 24) showed a spun list advantage in each experiment. A logistic regression analysis predicting the number of participants who showed the advantage from experiment was unsurprisingly not significant, Wald = 0.000, df = 1, B = 0.000, S.E. = 0.612, p = 1.000. A reduction in the number of people who detected structure in spun lists did not lead to a reduction in the number of people who showed the spun list advantage. This is in line with my previous tests on survey results, suggesting that detection of structure is not necessary to produce a spun list advantage.

A total of 32 participants showed the spun list advantage – 16 in each experiment. Treating the experiments as groups, I ran an independent-samples t-test on the difference in the magnitude of the spun list accuracy advantage in these 32 participants. The magnitude of the spun list advantage was lower in Experiment 5 (12.6%) than in Experiment 1 (21.8%), $t(30)$ = -1.956, difference = -0.092, one-sided p = .030. The disappearance of the spun list advantage is more consistent with the presence of interference in the item-to-item associations than with fewer people detecting the structure in spun lists.

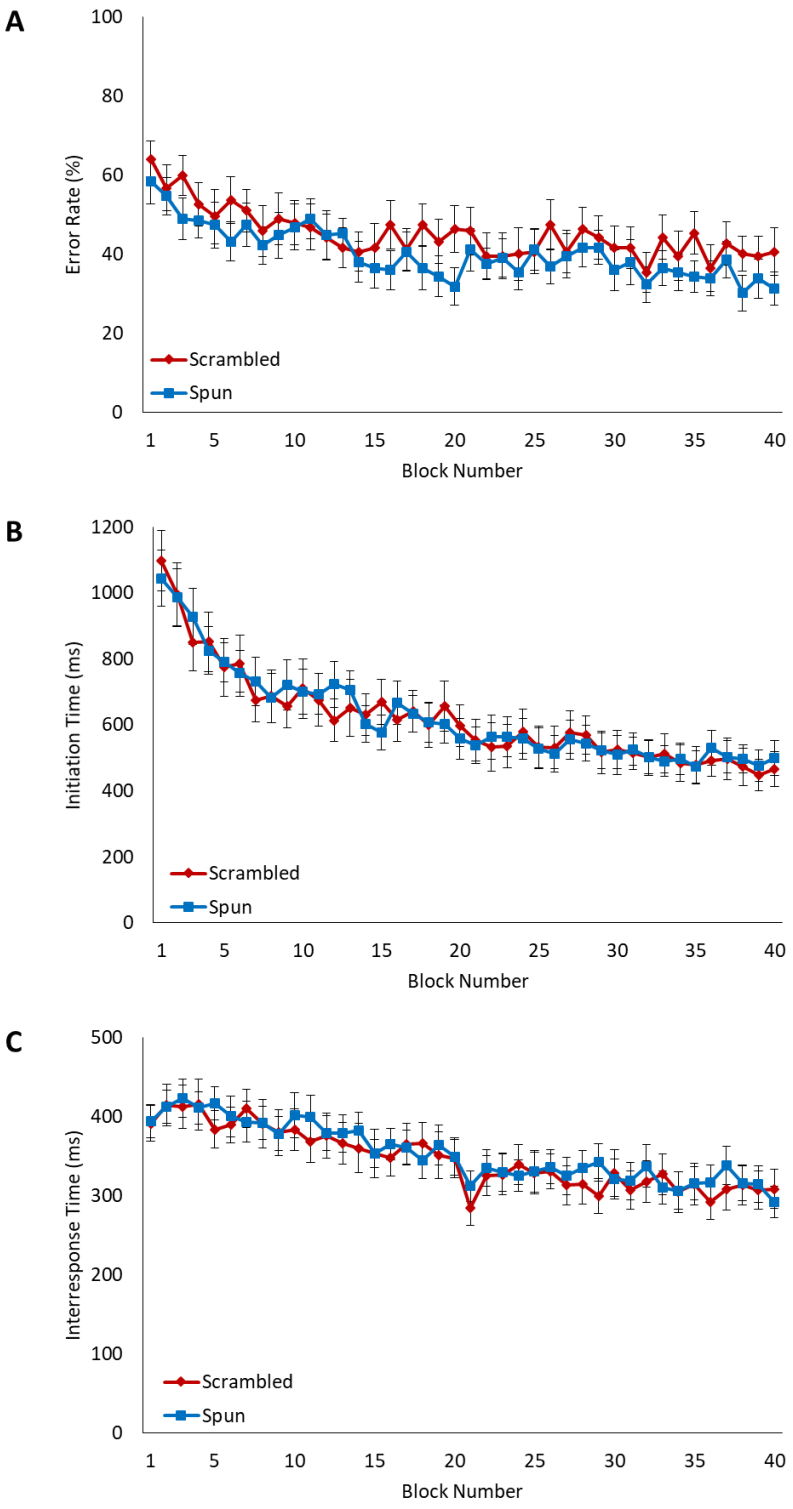

Figure 6. Experiment 5: Mean error rate (Panel A), IT (Panel B), and IRT (Panel C) for each list type as a function of the presentation number of the list (block number). The bars are standard errors of the means.

Table 14
Experiment 5: ANOVA and simple main effect analyses for training effects.

DV	F	dfs	MSE	p	ε	η_p^2
			Spun vs. Scrambled Performance			
Error Rate	2.312	1, 23	0.484	0.142	1.000	0.091
IT	0.016	1, 23	495,602	0.900	1.000	0.001
IRT	0.143	1, 23	182,474	0.708	1.000	0.006
			Learning (Spun)			
Error Rate	4.653	39, 897	0.021	< 0.001	0.886	0.168
IT	29.793	39, 897	16,181	< 0.001	0.364	0.564
IRT	9.727	39, 897	3,290	< 0.001	0.606	0.297
			Learning (Scrambled)			
Error Rate	4.150	39, 897	0.021	< 0.001	1.000	0.153
IT	30.200	39, 897	16,181	< 0.001	0.336	0.568
IRT	10.147	39, 897	3,290	< 0.001	0.578	0.306

The number of item-to-item associations that participants learned may also have played a role in reducing the magnitude of the spun list advantage. In Experiment 1, 6 items had consistent relative order, so 5 item-to-item associations supported performance in spun lists. In Experiment 5, there were only 4 spun letters and thus only 3 item-to-item associations to support performance. An additional experiment would be necessary to disentangle these possibilities, but that is beyond the scope of the current paper. The critical point is that detection did not cause the spun list advantage to disappear.

Table 15
Experiment 5: T-tests for first block performance and power function learning rate.

DV	t	df	M_{A-B}	SE_{A-B}	p	d	BF
			Spun vs. Scrambled First Block Performance				
Error Rate	-1.175	23	-0.057	0.049	0.252	-0.228	2.520 (N)
IT	-0.832	23	-51.865	62.309	0.414	-0.119	3.408 (N)
IRT	0.128	23	3.069	24.014	0.899	0.029	4.624 (N)
			Spun vs. Scrambled Learning Rate				
Error Rate	1.244	23	0.054	0.043	0.113	0.341	2.345 (N)
IT	0.264	23	0.026	0.100	0.397	0.073	4.512 (N)
IRT	0.743	23	0.069	0.0924	0.233	0.208	3.628 (N)

Power function fits. Average R^2 from fits to the data were .281, .410, and .432 for error rate, IT, and IRT, respectively. The average shared *b* parameters were 63.6, 949.4, and 264.4, respectively. The learning rate for spun lists was numerically but not significantly faster in error rate (spun: .217; scrambled: .164), IT (spun: .509; scrambled: .482), and IRT (spun: .489; scrambled: .420; Table 15). The tests on power function fits also indicate that the addition of interference in the associations led to no advantage for spun lists overall.

Strategies

Detecting the structure in spun lists has little bearing on the use of item-to-item associations, but strategy use might. Certain memorization strategies (e.g., rehearsal) may be more closely tied to item-to-item associations than others (e.g. grouping; Logie, 2018). Differences in the strategy used to recall the lists may produce differences in the likelihood of obtaining a spun list advantage or differences in the magnitude of the advantage.

Table 16
Descriptive statistics by strategy type.

		Error Rate		Spun List
Strategy	N	Spun	Scrambled	Advantage
None	6	0.328	0.452	0.124
Grouping	27	0.444	0.495	0.051
Sounding	38	0.392	0.492	0.100
Rehearsing	7	0.548	0.499	-0.048
Combination	66	0.324	0.429	0.105

Table 17
ANOVA on effects of strategy type.

DV	F	dfs	MSE	p	η_p^2
Spun Error Rate	2.589	1, 4	0.051	0.039	0.069
Scrambled Error Rate	0.570	1, 4	0.063	0.685	0.016
Spun Error Rate Advantage	1.177	1, 4	0.042	0.324	0.033

I combined the data from all 5 experiments and categorized the strategies reported in the post-experiment interviews for analysis. There were 5 categories: (1) none: reported no strategies; (2) grouping: reported focusing on a subset of the letters or parsing the letters into groups; (3) sounding: reported making sounds or syllables out of multiple letters; (4) rehearsing:

reported rehearsing the letter identities; and (5) combination: reported multiple strategies (most often a combination of grouping and sounding). Table 16 presents the number of people in each category, alongside average error rate for each list type and average error rate advantage for spun lists. I ran one-way ANOVA predicting these measures from the type of strategy used. Accuracy on spun lists, but not scrambled lists, varied with the type of strategy (Table 17). In spun lists, but not scrambled lists, there was structure for strategies to exploit. The spun list advantage differed among the strategies types, but not to a significant degree (Table 17). Strategy type was not predictive of the spun list advantage – it was observed in people with widely varying memorization strategies. This is additional evidence that the advantage reflects the use of item-to-item associations, and indicates that people are able to use a variety of strategies in concert with these associations.

Chapter Summary

In this chapter, I presented more rigorous analyses on the effects of detection and strategy use on the performance advantage for spun lists. In Experiment 5, I made it harder to detect structure in spun lists, but the additional detection difficulty did not affect the advantage. A variety of different strategies were reported over the presented experiments, but these strategies did not predict the spun list advantage either. Detection and strategy use cannot explain the spun list advantage.

I was able to make the spun list advantage go away in Experiment 5, and this seemed to be the result of the anchor items making relative order less consistent in the lists. This presents a boundary condition for when item-to-item associations are useful for serial memory retrieval. When there is no interference (as in Experiments 1-4), item-to-item associations help retrieval, but this benefit goes away when interference is introduced into these associations.

It is interesting that not everyone showed the spun list advantage. It demonstrates that even in ideal settings for item-to-item associations to be used, some people rely on other means of retrieval. At present, it is not clear what compels some to use these associations and others to not use them.

CHAPTER 5

SUMMARY AND DISCUSSION

I presented the results of 5 experiments that adapted the serial learning procedure of Lindsey and Logan (2019) to the domain of serial recall. These experiments demonstrated that people learn sequences more quickly when the relative order of items in those sequences is consistent, and that this learning transfers to new sequences if they share the same relative order. These findings are consistent with people using item-to-item associations to retrieve items in the sequence. The current experiments also demonstrated that these associations are formed between representations in memory, and that introducing interference into the associations makes them less useful routes of retrieval.

Meta-Analysis

Some of the presented experiments had issues with power, potentially calling into question the robustness of the spun list advantage. To put these concerns to rest, I conducted four meta-analyses – one each for error rate during training and test, and one each for IRT during training and test[3]. All five experiments were included in the training portion meta-analyses. Only Experiments 2, 3, and 4 were included in the test portion meta-analyses.

The meta-analytic method was taken from Borenstein et al. (2011). The difference between spun and scrambled lists was computed for each experiment and then transformed into Hedge's g (an unbiased estimate of the population effect size that is related to Cohen's d) using the equation:

$$g = \left(\frac{M_{diff}}{s_{within}}\right) * \left(1 - \frac{3}{4df - 1}\right)$$

[3] IT and IRT were not dissociated in Experiment 3. To include it in meta-analyses on timing data, I transformed the time taken measure from Experiment 3 into a measure of IRT by dividing time taken by 6 (the number of letters in the list). This assumes that 6 letter responses were given on each trial and that the initial letter response was not longer than the others. These assumptions are likely incorrect, but the transformation should give a good enough approximation of IRT to use in the meta-analyses.

A weighted average of these effect sizes was then computed. Each effect size received a weight equal to the inverse of its variance. The variance of g is given by the equation:

$$s_g{}^2 = \left(\frac{n_1 + n_2}{n_1 * n_2}\right) + \left(\frac{M_{diff}}{s_{within}}\right)^2 * \left(\frac{1}{2(n_1 + n_2)}\right)$$

A Z-statistic is then computed, in which the numerator is the weighted average effect size, and the denominator is the square root of one over the sum of the weights.

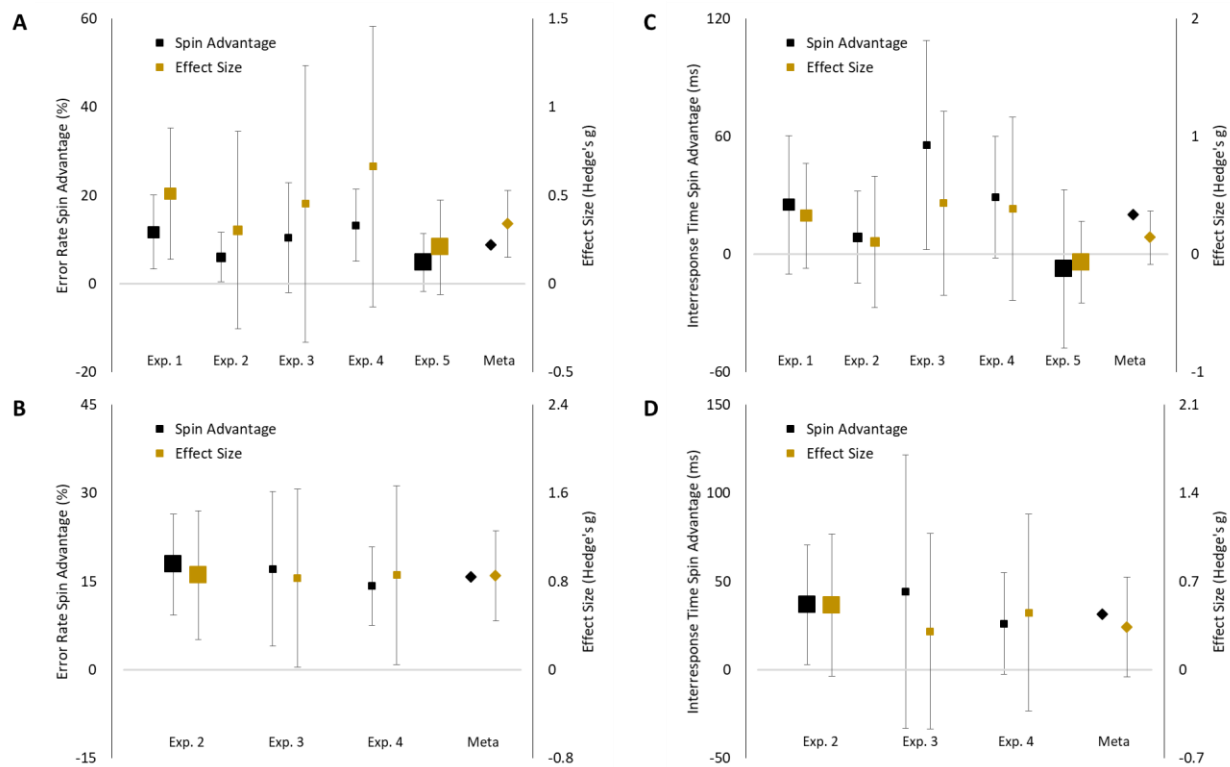

Figure 7. Spun list advantages (black data points) and their effect sizes (gold data points). Panel A and Panel B are error rate advantages for the training and test portions, respectively. Panel C and Panel D are timing advantages for the training and test portions, respectively. Square points are experiment data, and diamonds are the meta-analytic weighted averages. The sizes of the square points indicate how much weight they were given in the meta-analysis. The error bars are 95% confidence intervals; if the bar does not overlap with 0, there is a significant spun list advantage.

Table 18
Z-tests on Meta-analytic effect sizes

DV	Portion	Z	M_{diff}	g	SE_g	p
Error Rate	Training	3.480	0.087	0.339	0.097	< 0.001
	Test	4.088	0.158	0.851	0.208	< 0.001
IRT	Training	1.224	19.935	0.141	0.115	0.221
	Test	1.687	31.351	0.339	0.201	0.066

Figure 7 shows the difference between spun and scrambled lists in the training and test portions of each experiment. The black points are raw means, and the gold points are effect sizes. The error bars are 95% confidence intervals, so they indicate a significant test result if they do not overlap with zero. The tests on the raw means are equivalent to the list type main effect analyses presented earlier. There was a significant error rate advantage found in the training portions of Experiments 1, 2, and 4 and in the test portions of Experiments 2, 3, and 4. There was a significant IRT advantage in the training portion of Experiment 3 and the test portion of Experiment 2. The right-most gold points on the graphs show the tests on the meta-analytic effect sizes, and the results of the tests are presented in Table 18. This test was significant for error rate in the training portion and test portion, but not for IRT in either portion. The results of the meta-analyses are clear: Spun lists are remembered more accurately than scrambled lists in the training portion (8.7% accuracy advantage) and the test portion (15.8% accuracy advantage). The spun list advantage is real, and item-to-item associations really are used in serial recall.

Discussion

Most contemporary theories of serial memory retrieval assume that position-to-item associations (e.g., Henson, 1998) and graded activation (e.g., Page & Norris, 1998) are important features of the serial memory system. The conventional wisdom is that item-to-item associations are not used at all. The presented experiments show that item-to-item associations are also an important feature of the serial memory system.

The experiments shed light on some of the properties of the item-to-item associations used by the serial memory system. These associations seem to be learned quickly, have a maximum strength, and be susceptible to interference from other item-to-item associations. However, there are many properties of these associations that are currently unknown. We do not know if these associations are strictly local (formed between adjacent items) or if they can be

remote (formed between nonadjacent items). We do not know if the they are symmetric (i.e., the association between two items works in both directions; Asch & Ebenholtz, 1962) or direction-specific (the association only works in one direction; Wolford, 1971). We do not know the nature of the cue – they could be single item cues or compound cues (Murdock, 1995). Future research should address these questions to give a better understanding of how item-to-item associations are used by the serial memory system, and to inform how item-to-item associations should be instantiated in models of serial memory. Experiments 3 and 4 of Lindsey and Logan (2019) are examples of how some of these questions could be addressed.

It is reasonable to ask how the importance of item-to-item associations eluded researchers in this domain for so long. Logie (2018) summarizes the problem excellently: There is far too much emphasis on explaining the results of a task. Any one task might reveal just a sliver of the underlying cognitive system in which the researcher is interested. The standard serial recall task, for example, makes the position-to-item associations used by the serial memory system apparent, but not the item-to-item associations. Our understanding of memory for serial order benefits from the study of multiple task settings, because different tasks can illuminate different parts of the underlying system.

Although the current experiments make it apparent that item-to-item associations are a necessary component of theories of serial memory, they also show that these associations are not sufficient. For example, I invariably observed learning in scrambled lists that could not be attributed to consistency in relative order. This paper is not a call for a return to associative chaining theories. However, this is a call to abandon theories of serial memory – such as position-coding theories (Henson, 1998) or activation-based theories (Page & Norris, 1998) – that state or otherwise assume that previously retrieved items do not contribute to subsequent retrieval attempts. The evidence for item-to-item associations outweighs the appeal of parsimony.

The associations I observed were formed between representations in memory; they were not isolated to perception or action. Theories without item-to-item associations cannot be salvaged by assuming that these associations exist in some other stage of processing. Researchers will need to contend with the results I presented here when developing future models of serial memory – either by admitting that their scope is limited to a specific serial recall task, or by including item-to-item associations in their machinery. The most straightforward way to capture

the findings of the current paper would be to add item-to-item associations to an existing model of serial recall (see Burgess & Hitch, 1992 for an old hybrid model of serial memory that included position-to-item associations and item-to-item associations).

In free recall – a list memory task in which participants can report items in any order they wish – the contribution of previously retrieved items to subsequent retrieval attempts is less controversial than in serial recall (see Polyn, Norman, & Kahana, 2009; Raaijmakers & Shiffrin, 1980). Recent evidence (e.g., Ward, Tan, & Grenfell-Essam, 2010) suggests that there is a blurry boundary that separates the cognitive processes involved in free recall and serial recall. The results of the current experiments can be viewed similarly – people seem to use item-to-item associations in both tasks. The role item-to-item associations play in memory retrieval is likely a general one; keeping relative order during retrieval consistent with past experience makes retrieval easier (cf. Spillers & Unsworth, 2011; Tulving, 1962).

Item-to-item associations seem to be broadly important in tasks that require serial order. Typing and skill learning, for example, both seem to make use of item-to-item associations (Lindsey & Logan, 2019; Schuck et al. 2012). Speech and reading have historically been compared to serial memory (e.g., Bogaerts et al., 2016; Page & Norris, 2009; Vousden, Brown, & Harley, 2000), and it would be worthwhile to see if item-to-item associations are used in these domains as well. This research could reveal a general mechanism of serial order that is used in multiple domains (e.g., a response-driven context retrieval mechanism; Logan, 2018; Polyn, Norman, & Kahana, 2009).

Conclusion

The mechanism of serial memory is more complex than the conventional position-based theories depict. People seem to have multiple cognitive tools that they can bring to bear to solve the problem of serial order, and the tools they use depend on the task demands. When relative order is consistent, item-to-item associations are one of the tools that get used.

REFERENCES

Asch, S. E., & Ebenholtz, S. M. (1962). The principle of associative symmetry. *Proceedings of the American Philosophical Society, 106*, 135-163.

Bogaerts, L., Szmalec, A., De Maeyer, M., Page, M. P., & Duyck, W. (2016). The involvement of long-term serial-order memory in reading development: A longitudinal study. *Journal of Experimental Child Psychology, 145*, 139-156.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis.* John Wiley & Sons.

Brown, G. D., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107*, 127-181.

Burgess, N., & Hitch, G. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language, 31*, 429-460.

Caplan, J. B. (2015). Order-memory and association-memory. *Canadian Journal of Experimental Psychology, 69*, 221-232.

Crump, M. J. C., & Logan, G. D. (2013). Prevention and correction in post-error performance: An ounce of prevention, a pound of cure. *Journal of Experimental Psychology: General, 142*, 692–709.

Ebbinghaus, H. (1885). *Über das gedächtnis*. Leipzig, Germany: Duncker and Humblot.

Ebenholtz, S. M. (1963). Serial learning: Position learning and sequential associations. *Journal of Experimental Psychology, 66*, 353-362.

Farrell, S. (2012). Temporal clustering and sequencing in working memory and episodic memory. *Psychological Review, 119*, 223– 271.

Farrell, S., Hurlstone, M. J., & Lewandowsky, S. (2013). Sequential dependencies in recall of sequences: Filling in the blanks. *Memory & Cognition, 41*, 938-952.

Fischer-Baum, S., & McCloskey, M. (2015). Representation of item position in immediate serial recall: Evidence from intrusion errors. *Journal of Experimental Psychology, 41*, 1426–1446.

Henson, R. N. A. (1996). *Short-term memory for serial order* (Unpublished doctoral dissertation, University of Cambridge).

Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36*, 73-137.

Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *The Quarterly Journal of Experimental Psychology, 49A*, 80-115.

Hitch, G. J., Fastame, M. C., & Flude, B. (2005). How is the serial order of a verbal sequence coded? Some comparisons between models. *Memory, 13*, 247-258.

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin, 140*, 339-373.

Kahana, M. J., Mollison, M. V., & Addis, K. M. (2010). Positional cues in serial learning: The spin-list technique. *Memory & Cognition, 38*, 92-101.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. (pp. 112-146). New York: Wiley.

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. *Psychology of Learning and Motivation, 49*, 1-48.

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review, 96*, 25-57.

Lindsey, D. R. B., & Logan, G. D. (2019). Item-to-item associations in typing: Evidence from spin list sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*, 397-416.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492-527.

Logan, G. D. (2018). Automatic control: How experts act without thinking. *Psychological Review, 125*, 453-485.

Logan, G. D., & Crump, M. J. C. (2010). Cognitive illusions of authorship reveal hierarchical error detection in skilled typists. *Science, 330*, 683–686.

Logan, G. D., Ulrich, J. E., & Lindsey, D. R. B. (2016). Different (key)strokes for different folks: How standard and nonstandard typists balance Fitts' law and Hick's law. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 2084-2102.

Logie, R. (2018). Human cognition: Common principles and individual variation. *Journal of Applied Research in Memory and Cognition, 7*, 471-486.

Logie, R. H., Saito, S., Morita, A., Varma, S., & Norris, D. (2016). Recalling visual serial order for verbal sequences. *Memory & Cognition, 44*, 590-607.

MathWorks. [MATLAB R2018a]. (2018).

Murdock, B. B. (1995). Developing TODAM: Three models for serial-order information. *Memory & Cognition, 23*, 631-645.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Osth, A. F., & Dennis, S. (2015). The fill-in effect in serial recall can be obscured by omission errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 1447-1454.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*, 761–781.

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*, 3737-3753.

Polyn, S. M., Norman, K. A., and Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116*, 129–156.

Psychology Software Tools, Inc. [E-Prime 2.0]. (2012).

Raaijmakers, J. G. W., & Shiffrin, R. (1980). SAM: A theory of probabilistic search of associative memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 207–262). New York, NY: Academic Press.

Raskin, E., & Cook, S. A. (1937). The strength and direction of associations formed in the learning of nonsense syllables. *Journal of Experimental Psychology, 20*, 381-395.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225-237.

Scarborough, D. L. (1972). Memory for brief visual displays of symbols. *Cognitive Psychology, 3*, 408–429

Schuck, N. W., Gaschler, R., Keisler, A., & Frensch, P. A. (2012). Position-item associations play a role in the acquisition of order knowledge in an implicit serial reaction time task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 440-456.

Solway, A., Murdock, B. B., & Kahana, M. J. (2012). Positional and temporal clustering in serial order memory. *Memory & Cognition, 40*, 177-190.

Spillers, G.J. & Unsworth, N. (2011). Variation in working memory capacity and temporal-contextual retrieval from episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1532-1539.

Snyder, K. M., Ashitaka, Y., Shimada, H., Ulrich, J. E., & Logan, G. D. (2014). What skilled typists don't know about the QWERTY keyboard. *Attention, Perception & Psychophysics, 76*, 162-171.

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review, 69*, 344-354.

Vousden, J. I., Brown, G. D., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology, 41*, 101-175.

Ward, G., Tan, L., & Grenfell-Essam, R. (2010). Examining the relationship between free recall and immediate serial recall: the effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1207-1241.

Wolford, G. (1971). Function of distinct associations for paired-associate performance. *Psychological Review, 78*, 303-313.

Yamaguchi, M., & Logan, G. D. (2016). Pushing typists back on the learning curve: Memory chunking in the hierarchical control of skilled typewriting. *Journal of Experimental Psychology: Learning, Memory and Cognition, 42*, 1919-1936.