



Learning and production of movement sequences: Behavioral, neurophysiological, and modeling perspectives

Bradley J. Rhodes^a, Daniel Bullock^{a,*}, Willem B. Verwey^b,
Bruno B. Averbeck^c, Michael P.A. Page^d

^a *Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, MA 02215, USA*

^b *Department of Psychonomics and Human Performance, University of Twente, 7500 AE Enschede, The Netherlands*

^c *Center for Visual Science, Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627-0270, USA*

^d *Psychology Department, University of Hertfordshire, Hatfield AL10 9AB, UK*

Abstract

A wave of recent behavioral studies has generated a new wealth of parametric observations about serial order behavior. What was a trickle of neurophysiological studies has grown to a steady stream of probes of neural sites and mechanisms underlying sequential behavior. Moreover, simulation models of serial behavior generation have begun to open a channel to link cellular dynamics with cognitive and behavioral dynamics. Here we review major results from prominent sequence learning and performance tasks, namely immediate serial recall, typing, $2 \times N$, discrete sequence production, and serial reaction time. These tasks populate a continuum from higher to lower degrees of internal control of sequential organization and probe important contemporary issues such as the nature of working-memory representations for sequential behavior, and the development and role of chunks in hierarchical control. The main movement classes reviewed are speech and keypressing, both involving small amplitude movements amenable to parametric study. A synopsis of serial order models, vis-à-vis major

* Corresponding author. Tel.: +1 617 353 9486; fax: +1 617 353 7755.
E-mail address: danb@cns.bu.edu (D. Bullock).

empirical findings leads to a focus on competitive queuing (CQ) models. Recently, the many behavioral predictive successes of CQ models have been complemented by successful prediction of distinctively patterned electrophysiological recordings. In lateral prefrontal cortex, parallel activation dynamics of multiple neural ensembles strikingly matches the parallel dynamics predicted by CQ theory. An extended CQ simulation model – the N-STREAMS neural network model – exemplifies ongoing attempts to accommodate a broad range of both behavioral and neurobiological data within a CQ-consistent theory.

© 2004 Published by Elsevier B.V.

PsycINFO classification: 2330; 2343; 2420; 2500; 4160

Keywords: Serial learning; Motor performance; Cognitive processes; Working memory; Computer simulations; Neural networks; Competitive queuing; Primacy; Prefrontal cortex; Cerebellum; Basal ganglia

1. Introduction: A brief history of serial order

Thinking about movement sequences has a long history in behavioral science. Pavlov and other early observers (for a review, see e.g., Adams, 1984) noted that sequences may arise if feedback caused by generating one response triggers the next one. This mechanism has been called stimulus–response reflex chaining or simply *response chaining* (Bain, 1868; James, 1890). That movement sequences can be performed in the absence of sensory feedback argues against the sufficiency of response chaining. Such cases suggest centrally programmed command sequences that are “structured before the movement begins and allow . . . the entire sequence to be carried out uninfluenced by peripheral feedback” (Keele, 1968, p. 387). Precursors of this centralist view date back at least as far as Helmholtz (1866), James (1890), Woodworth (1899), and Lashley (1917, 1951).

If sequence production is normally insensitive to removal of expected sensory feedback, then sequences are executed under open-loop control. A testable version of open-loop sequence control was formalized by Henry & Rogers (1960). Their ‘memory drum’ model started with the observation that simple RT increased as a function of movement complexity, an effect reported much earlier (e.g., Freeman, 1907). The memory drum model explained these results by assuming that innate and learned neuromotor coordination patterns are stored, and become accessible for production, via a mechanism abstractly similar to the rotating drums that memory psychologists then used to automate presentations of stimuli (inscribed on the curved surface) to subjects, at a rate dependent on variables such as drum radius and shaft rotation speed. They proposed that the net drum operation rate slows as the sequence to be recalled becomes more complex. Although a ‘complexity’ metric was not defined in 1960, Henry (1980) wrote that he and Rogers intended the dictionary definition: “That is complex which is made up of a number of connected parts” (p. 164). Thus, a more complex response would have a larger number of connected parts than a less complex response. Although the empirical bases of this theory came from *simple RT* (sRT) tasks, Henry argued that the theory applied to *choice RT* (cRT) tasks as well. The complexity effect has been replicated many times in various types of tasks, including

some that tested alternative explanations for the effect (Anson, 1982; Christina, Fischman, Vercruyssen, & Anson, 1982). However, one prediction of the memory-drum (or phonograph) metaphor – that performance rate for all sequence elements should *uniformly* slow or quicken – has been falsified in multiple studies of human performance (Heuer, 1988; MacKenzie & van Eerd, 1990; Verwey, 2003b).

A distinct line of research has examined skilled performance. Perceptual-motor skills exemplified by typing and various sports include complex movement sequences. A recurring proposal is that complex movement sequences are controlled *hierarchically* (e.g., Book, 1908; Miller, Galanter, & Pribram, 1960). Hierarchical control is often depicted with tree-like branching structures. Skills involve units of behavior, and each high level unit is subdivided into lower-level units that involve smaller and more explicitly defined units, until, at the lowest level, limbs or muscles are specified. The higher levels deal with longer term consequences, and lower levels consider short-term details of individual movements. Level of control has been associated with modifiability: “If the ‘vital’ centers of the lowest levels were not strongly organized at birth, life would not be possible; if the centers on the highest levels (‘mental centers’) were not little organized and therefore very modifiable we could only with difficulty and imperfectly adjust ourselves to the circumstances and should make few acquisitions” (Taylor, 1932, p. 437).

Thus, hierarchical control would combine autonomous functions at low levels with the possibility of learning new operations at higher control levels. By the late 1970s, this general idea was posited in testable form (e.g., Reason, 1979; Rosenbaum, Kenny, & Derr, 1983; Sternberg, Monsell, Knoll, & Wright, 1978, 1980). Since then, some connectionist theorists have argued that it may be possible to explain the data indicating hierarchically organized behavior without accepting the inference of a genuine hierarchical controller (e.g., Elman, 1990; for critiques, see Page, 1994; MacWhinney, 2003). This is one of the key theoretical issues to which we will return following a survey of recent data. These data, which include not only parametric behavioral effects but also a flood of recent neurobiological observations, establish an extremely challenging set of criteria that must be met by models of serial order. It has become clear that comprehensive serial order models of the future will have to include multiple bases for sequence storage and learning, and that each basis will involve distinctive representational resources. There is now compelling evidence that part of such a model should include a component that can represent and store both novel and familiar sequences as parallel activation patterns without using any direct associative links between successive items in the represented sequence. Such evidence violates the predictions of many classical and recent connectionist models, but supports the predictions of one class of ‘hierarchy friendly’ connectionist models, namely competitive queuing (CQ) models (Grossberg, 1978a, 1978b; Houghton, 1990).

2. Sequence learning and performance research: Major active paradigms

We now review major data constraints pertinent to delineating mechanisms responsible for proficient representation and execution of movement sequences,

especially short sequences that are known before movement is initiated. Most conceptual models have been developed in the context of a particular task. There are distinct schemes (conceptual models) for handwriting, typing, speech production, and musical performance. These schemes have a mutual resemblance but are often too sketchy to allow definition of common and distinctive features. That these schemes all describe examples of the same phenomenon – skilled human motor performance – justifies the quest for a generic formal (e.g., neural network) model of human motor behavior that encompasses the basic principles of skilled behavior.

The reviewed paradigms probe issues such as internal sequence preparation, hierarchical control and practice effects. All paradigms consist of tasks which employ small amplitude movement elements that can be produced rapidly (such as key-presses and uttering phonemes), because sequences composed from movements with prolonged durations can conceal preparation for forthcoming sequence elements that may occur during execution of prior elements (e.g., Verwey, 1996, 2001). Most reviewed tasks also involve moderate to extensive practice. Such practice leads to skilled performance in which sequences are based neither on deliberative choice of individual acts nor on the stimulus guidance that characterizes assembly of unfamiliar sequences.

2.1. Tasks and data treated

The tasks (Table 1) to be reviewed populate an *internal control* continuum that ranges from tasks that strictly require internal (e.g., long-term- or working-memory-guided) sequence control (ISR, $2 \times N$, PSFD, plus simple and choice RT tasks with sequences for responses), through tasks that strongly encourage but do not strictly require internal sequence control (e.g., DSP), to tasks that allow but do not encourage internal sequence control (e.g., SRT). Chronometric and/or error pattern analyses, as available, will be summarized to identify robust trends to constrain models of sequence representation and production.

2.1.1. Immediate serial recall (ISR)

In its canonical form, the ISR (immediate serial recall) task involves the presentation of a list of familiar items (e.g., digits, letters, words) that the participant is asked to recall in the correct order. The participant's response exemplifies a planned movement sequence. In spoken ISR, the necessary coordination of the articulators during the utterance is one of the most complex motor acts that we routinely attempt. That it rarely impresses us as virtuosic, testifies to the massive amount of everyday speech practice from which such performance benefits.

Later we review evidence that the standard ISR task draws heavily on speech-based processes. First we outline the general form of ISR data, drawing attention to ways in which they differ in emphasis from other sequential task data. One principal difference is that most ISR data analyses treat patterns of errors. Since early studies of Conrad and Brown, error analysis has been a 'royal road' to the study of memory (Henson, 1996). Indeed, the large majority of modern models of short-term serial memory are tested against error data as opposed to timing data. Yet

Table 1

List of serial order behavioral paradigms reviewed, plus other abbreviations used

Abbreviation	Expansion
<i>Serial order behavioral paradigms</i>	
$2 \times N$	$2 \times N$ task – where N is typically 5 or 10
cRT	choice reaction time (with response being a sequence)
DSP	Discrete sequence production
ISR	Immediate serial recall
PSFD	Prescribed sequence figure drawing
sRT	simple reaction time (with response being a sequence)
SRT	Serial reaction time
Typing	Typing – primarily with short sequences
<i>Paradigm parameters/measures</i>	
ERP	Event-related potentials
IRI	Inter-response interval
LRP	Lateralized readiness potential
RSI	Response to stimulus interval
<i>Effects from data</i>	
SLEL	Sequence length effect on latency
SLER	Sequence length effect on (mean production) rate
WLE	Word-length effect
<i>Brain activity investigation techniques</i>	
EEG	Electroencephalography
fMRI	Functional magnetic resonance imaging
PET	Positron emission tomography
TMS	Transcranial magnetic stimulation
<i>Models/networks</i>	
CQ	Competitive queuing
HED	Hierarchical editor
RNN	Recurrent neural network
<i>Cortical regions</i>	
DLPFC	Dorso-lateral prefrontal cortex
IPS	Intra-parietal sulcus
SMA	Supplementary motor area

theorists and modelers have not neglected ideas relating to the passage of time, and a fierce debate continues regarding the effects of passing time on ISR performance. This debate concerns the effects both of short, filled delays on ISR performance and of other within-recall delays brought about by differential articulation times for the verbal materials. Timing phenomena are more often considered for their effects on the number and pattern of errors than as explicit targets for modeling.

A recurring target for ISR modeling is the *serial position curve* produced by plotting the number of serial-recall errors against output position. An error is any failure to recall, in a given output position, the item that was presented in the corresponding position in the stimulus list. Error types include transpositions of list-items from other positions in the same list and the omission of any response at a position (which participants are asked to indicate with the word “blank” during spoken recall, or

with a dash in written recall). For ISR of lists of approximately *span length* – span being the length of a list that a participant can recall correctly half the time – the serial-recall curve has a very well-established form, that illustrated in Fig. 1 (from an ISR experiment involving six visually presented items). It is often glossed as *bow-shaped*, by analogy with curves from *free-recall* experiments, in which response order is unconstrained. However, ‘bow-shaped’ suggests a symmetry in the curve that is rarely present. A better summary is that errors increase approximately linearly across output positions, with a dip at the list-final position. The advantage for early items is normally dubbed a *primacy effect*, and the improved performance for the final item (occasionally items) is dubbed a *recency effect*.

The characteristic form of the serial position curve proved surprisingly difficult to simulate in a number of sophisticated models of ISR. For example, the influential connectionist model of Burgess & Hitch (1992) had considerable trouble capturing this aspect of the data. An observation made by Page and colleagues (Henson, Norris, Page, & Baddeley, 1996; Norris, Page, & Baddeley, 1994; Page & Norris, 1998) helped to resolve this problem. They noted a dominant transposition error pattern, which they termed *fill-in*: an item’s being recalled early, say at position n , is most often immediately succeeded by ‘fill-in’ of the item that should have appeared at position n . Such fill-in is contrary to the prediction of models that represent serial order via association links between prior states and next items. Fill-in implies that if the response to the letter-list “RXKHZB” begins “RXH...” then recall is much more likely to continue with a “K” than with, say, a “Z”, whereas ‘chaining’ models incorrectly predict that the “H” would tend to link onwards to its successor “Z” rather than back to its predecessor “K”. Lack of fill-in mitigated against the bias toward adjacent item transpositions that was found in the data and turned out to be a crucial problem that prevented most models that preceded the primacy model of ISR (Henson et al., 1996; Page & Norris, 1998) from accurately modeling the serial position curve.

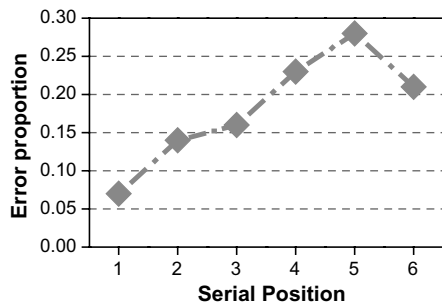


Fig. 1. A schematic serial position curve for immediate serial recall of a span-length list. The plot is typical of data reported by Henson et al. (1996) that strongly informed the modeling of Page and Norris (1998). The approximately linear increase in error rate across serial position, except for a dip below the trend line in the list-final position, is apparent.

Factors other than serial position also influence ISR errors. Substantial evidence suggests that speech processes, including operations of a short-term, speech-based store, are crucial for ISR performance under normal circumstances. Much of this evidence was collected with reference to the working-memory framework established by [Baddeley & Hitch \(1974\)](#) and later developed by [Baddeley \(1986\)](#). The first result indicative of the use of a speech-based store was termed the *phonological similarity effect* (e.g., [Baddeley, 1968](#); [Conrad, 1964](#)). The recall of lists of rhyming items (e.g., the list of letter names “GCBTPV”) is reliably poorer than the recall of lists of non-rhyming items (e.g., “HRQXBL”). This robust effect is seen even if participants are shown the stimulus lists visually, provided that they are not forced to engage in concurrent articulation during the visual presentation. Such concurrent articulation normally takes the form of repeatedly uttering an irrelevant word (e.g., “blah blah blah”) and is taken to block the process by which the visual stimulus is recoded into the speech-based store. With no chance to recode, participants are assumed to perform from a visual, or perhaps episodic, memory within which phonological similarity plays no role. Interestingly, concurrent articulation has no effect on the phonological similarity effect when auditory presentation is used. This is taken to indicate that auditory materials’ access to the speech-based store is ‘direct’.

A similar input-type-by-task interaction effect seen in ISR is the *irrelevant sound effect*. Performance in ISR is disrupted in the presence of irrelevant background sound of a particular type ([Colle & Welsh, 1976](#); [Salamé & Baddeley, 1982](#)). If the irrelevant sound is ‘white noise’, there is no disruption, but an irrelevant stimulus that changes its state ([Jones, Madden, & Miles, 1992](#)) does disrupt ISR. Irrelevant speech sounds (even in a language unfamiliar to the participant) produce strong effects. The ‘unattended speech effect’ was later renamed the ‘irrelevant sound effect’, following findings that speech is neither necessary nor sufficient. [Jones & Macken \(1993\)](#) first showed that a changing sequence of tones suffices to disrupt ISR, whereas an unchanging irrelevant speech stimulus, such as repeated utterance of the same token, only weakly disrupts ISR. That concurrent articulation abolishes the irrelevant sound effect for visual but not auditory presentation of ISR materials, combines with the fact that disruption is strongest when the irrelevant stimulus changes in a speech-like manner to give qualified support to the hypothesis that a speech-based store underlies standard ISR.

Two further factors affecting errors in ISR are delayed production and the length of list-items, notably word-length. Even short, filled delays intervening between list presentation and the recall attempt can cause drastic drops in serial-recall performance (e.g., [Conrad, 1958](#)). For this result to accrue, the brief retention interval must be filled with some task that prevents subvocal rehearsal of the list: performance can survive practically indefinite delays in which participants are able to engage in such rehearsal. The need to fill the retention interval introduces some ambiguity as to the cause of the drop in performance with delay. Is it due to the passage of time as such? Or is it due to some interference from the activity with which the retention interval is necessarily filled? Or is it both? Because this 40 year debate continues to rage, it is unrealistic to reprise it here, but time-related memory decay – an explanation

favored by [Baddeley & Hitch \(1974\)](#) – remains a viable candidate for explaining at least part of the effect.

The *word-length effect* (WLE) is usually grouped with the effect of delay because both might be attributable to memory decay or memory interference. The basic result is that lists of long words (e.g., five-syllable words like “university”) are recalled worse than lists of short words (e.g., one-syllable words like “chair”) when correct serial order is required. Thus, the WLE contradicts [Miller’s \(1956\)](#) view that span is about seven chunks, whatever those chunks might be. Originally, two possible explanations of the WLE presented themselves. First, that the rehearsal of long words is more difficult than rehearsal of short words, so that when participants came to recall, their rehearsal would have been more recent in the case of the short-word lists. The consequences of a more recent rehearsal, in a system in which memory is decaying over time, would be seen in improved performance. This view was later extended to incorporate time delays during output. It was noted that simply uttering the early words in the recall of a long-word list would delay the recall of later words, relative to the later words in responses to short-word lists (e.g., [Cowan, 1994](#)). In both the original and the extended account, the passage of time is held to play a crucial role. The second type of explanation involved an appeal to some capacity limitation in short-term memory. The long words, with their increased number of phonemes or syllables, simply used up more of this capacity, resulting in poorer recall.

In order to distinguish between these explanations, [Baddeley, Thomson, & Buchanan \(1975\)](#) tested recall of words of different lengths, with word-length measured in terms of articulation time rather than in terms of phonemic or syllabic length. They thus used short words like “wicket” and “bishop” as opposed to long words like “voodoo” and “zygote”. They found a reliable WLE that they attributed to delay and decay rather than to a capacity limitation. In its reliance on the time to articulate, the WLE thus fits with the other evidence relating ISR to speech-based processes.

The [Baddeley et al. \(1975\)](#) result has been empirically challenged numerous times, with some claiming that although the result replicates with the precise set of words used by [Baddeley et al.](#), it fails to generalize to other word sets ([Caplan, Rochon, & Waters, 1992](#); [Lovatt, Avons, & Masterson, 2000, 2002](#)). The issues of how delay and word-length causally affect ISR performance remain subjects of hot dispute, but few would now deny a link with processes relating to speech output.

2.1.2. Typing

Typing involves the production and learning of relatively discrete sequences of items. This section will focus on typing studies in which the sequences are relatively short. Nonetheless, a brief consideration of more continuous transcription typing data permits the identification of some general phenomena which both concur and conflict with the discrete typing data in focus here.

[Sternberg et al. \(1978\)](#) asked well-trained typists to produce short sequences of keystrokes as rapidly as possible from memory. To focus on the production aspects of the task, the sequences were presented at the start of a trial and subjects were

given ample time to internally prepare their response. There was also a count down to the imperative/GO signal which cued initiation of the prepared response. Although the subjects were competent typists, the sequences presented were not well practiced words and were not presented repeatedly enough to induce practice effects. Sequence length was the primary independent variable, although some sequences required performance with one hand while others required alternation between hands. The major dependent variables were: latency (or reaction time, RT), measured as the delay between presentation of the GO signal and the first keypress; and inter-response intervals (IRIs), the latencies between any two consecutive keypresses.

Sternberg et al. (1978) found a number of consistent effects. First was a *sequence length effect on latency* (SLEL): latency increased as a linear function of sequence length. Second was a *sequence length effect on rate* (SLER): mean IRI (an inverse measure of production rate across the entire sequence) also increased as a function of sequence length. Third, the ratio of latency to mean IRI – the *ratio effect* – was much greater than one. Fourth, there were *serial position effects* on individual IRIs. For a given sequence length, some IRIs were longer than others, but the longer IRIs occurred in different positions depending on sequence length. Fifth, the IRI preceding performance of the final item in a sequence was shorter than that which preceded it – initiation of the final item was fast (see Fig. 2). This pattern of results was

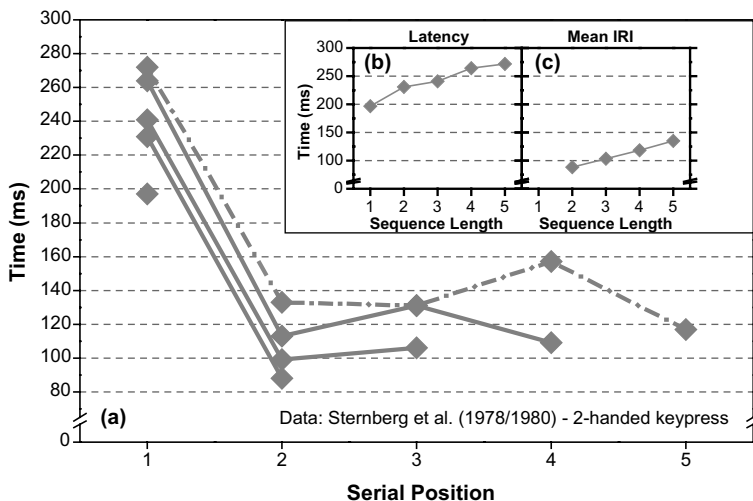


Fig. 2. Alternating hand typing results from Sternberg et al. (1978). (a) Reaction time and inter-response intervals (IRIs) for each sequence length, illustrating the time course of performance. The 5-item sequence data is presented with a dashed line to resolve potential ambiguity due to the intersection of the lines for the 4- and 5-item sequences at the third item; (b) Sequence length effect on latency (SLEL); (c) Sequence length effect on mean IRI (the inverse of production rate) – (SLER). The grey color signifies performance considered to be at an early level of practice. Note that the ratio of the latency (sequence start time) to the mean IRI is much greater than 1.0. This long-established ratio effect is a chronometric signature of preparation of an entire sequence before initiation, which is possible even for novel sequences in this paradigm. Using such whole-sequence preparation was recently referred to as the ‘collective’ planning strategy in Conway and Christiansen (2001).

obtained for both single and alternating hand conditions and for similar speech production tasks using one or two syllable words. Error rates were very low, and no analysis of types or frequency of errors was reported. A subsequent study, where sequence lengths were increased to 6 items, reported a large increase in error rate for 6-item sequences (Sternberg, Wright, Knoll, & Monsell, 1980). The chronometric patterns have generally been replicated in a variety of investigations; however, they do vary under some circumstances – such as after brain injury or following practice of the sequences.

Effects of various brain lesions have been investigated using the Sternberg et al. (1978) paradigm. Rafal, Inhoff, Friedman, & Bernstein (1987) found that Parkinsonian patients exhibited the same basic pattern of results (as those of unimpaired control subjects), but displayed slower overall execution rates. Damage to the basal ganglia apparently did not change the basic temporal aspects of sequence performance as identified by Sternberg et al. (1978). When patients with bilateral cerebellar lesions were tested, however, the pattern of results changed radically for moderately, but not mildly, impaired subjects (Inhoff, Diener, Rafal, & Ivry, 1989). *Moderately affected cerebellar patients exhibited no SLEL, and IRIs were not significantly shorter than latencies*. Replication with unilaterally affected cerebellar patients revealed the same set of abnormal effects only on the affected (ipsilesional) side of moderately impaired subjects. Further research concluded that damage to the *lateral* cerebellum, sufficient to produce moderate clinical impairment, resulted in the altered timing patterns (Inhoff & Rafal, 1990), whereas damage to *medial* cerebellum did not produce timing patterns distinct from those found by Sternberg et al. (1978). This dissociation is revealing because the lateral (but not the medial) cerebellum is reciprocally connected with the frontal cortex.

In addition to the cerebellar-deficit-related changes in temporal patterning, there are reports of practice-induced timing changes in sequential keypressing tasks. Although participants in Sternberg et al. (1978) were practiced at the task, they were not highly practiced on the individual sequences. Further studies have shown that, depending on the task, the SLEL (sequence length effect on latency) disappears after moderate to extensive levels of practice (more than about 300 practice trials, e.g., Klapp, 1995; Verwey, 1999, fig. 3). More extensive practice (of around 2500 trials) can eliminate the serial position effects on IRIs (e.g., Verwey, 1996). The high ratio of latency to IRI, the SLER (sequence length effect on rate), and faster final IRI were reported to survive even this extreme level of practice.

A total of 29 transcription typing phenomena were identified in a comprehensive review by Salthouse (1986). The first relevant phenomenon is that the rates of typing for random orders of real words and for meaningful text do not differ markedly. On the other hand, the rate of typing is slowed as the letters to be typed approach random sequences. Increased practice with the same sequences reduces the randomness, and production rate increases. This pattern is consistent with the keypress studies of Sternberg et al. (1978), Klapp (1995), and Verwey (1996). A third relevant phenomenon is that restricted preview severely impairs typing rate. This prevents typists from preparing future responses in advance – a situation explicitly avoided in the Sternberg et al. (1978) paradigm but probed in Verwey's studies (as described below).

Fourth, sequences produced with alternating hands have a higher rate than those performed with fingers of the same hand – as was the case in Sternberg et al. (1978) (though this may reverse with extensive practice, Verwey & Clegg, *in press*). Fifth, pairs of letters that occur more frequently in normal text are typed more rapidly than less frequently occurring pairs. This held true even when the type of transition between fingers was controlled. It is likely to be the result of extensive practice with high frequency digrams, consistent with the learning-driven changes outlined above. Sixth, greater practice driven changes occur for digrams typed with two different fingers than those typed with the same finger – learning to overlap and coordinate movements for consecutive keystrokes (possible only when different effectors are involved) is obviously an important aspect of learning.

One phenomenon identified by Salthouse (1986) is apparently in contradiction with two of the main Sternberg et al. (1978) effects. Taken as a whole, the transcription typing data indicate there is no systematic SLEL (when latency is measured between depression of the space bar and depression of the first letter of a word) or SLER. Salthouse attributed the discrepancies with the Sternberg results to the difference between ‘discontinuous or burst typing’ (as in Sternberg et al.) and continuous (transcription) typing. On the other hand, the ratio effect of Sternberg et al. (1978) – where latency is much greater than the subsequent IRIs – is a robust phenomenon of transcription typing (Salthouse, 1986). Also, increases in typing skill decrease the variability of IRIs (Salthouse, 1986), consistent with the loss of a serial position effect on IRI after extended practice (Verwey, 1996).

Regarding errors, Salthouse (1986) reported the following: substitutions are the most prevalent type of error among novice typists but intrusions (where an extra character is inserted) and omissions become more prevalent in advanced typists; transpositions (where the order of two adjacent characters is reversed) represent a small percentage of errors. Many substitution errors involve adjacent keys. This suggests that they are not central sequencing errors but arise from misdirected movements, especially among novices. Extremely short IRIs occur in association with many intrusion errors, which often result from ‘double keypresses’ pursuant to a single (slightly misdirected) finger depression or from failure to properly deactivate a prior keystroke causing an unwanted repetition. A longer IRI often follows omission errors, which may arise when finger movement is insufficient to depress the relevant key (especially so for difficult to reach keys); most transposition errors are cross-hand rather than within-hand. Thus output mechanics, rather than central processes, seem to dominate transcription typing, which by definition does not tax working memory for arbitrarily novel sequences.

2.1.3. The $2 \times N$ task

The ‘ $2 \times N$ ’ task (Hikosaka, Rand, Miyachi, & Miyashita, 1995), popular in non-human primate studies, begins with a sequence discovery phase, during which subjects use trial and error to discover the correct sequence. After discovery, continued practice consolidates learning and performance. Subjects face a 4×4 square matrix of buttons. A trial consists of an initial home key press followed by up to 5 (for monkeys) or 10 (for humans) consecutive 2-item ordering problems. For each problem,

the 4×4 array is presented with only two of the buttons lit. The subject must press these two buttons in the correct sequence – initially unknown to the subject. Any error aborts the trial and causes reset back to the start of the $2 \times N$ sequence. Upon correct 2-item ordering, the lights are extinguished (for as long as a second in some studies) and then two new buttons are illuminated – the next problem. Each such 2-item ordering problem is called a *set*, and the entire sequence of N (5 or 10) sets is called a *hyperset*. Typically, hypersets are represented to subjects until they successfully complete the entire hyperset a predetermined number of times. These errorless completions do not have to occur consecutively – they accumulate within a block. When the required number of successful trials has been performed, the block is complete and there is a switch to a different hyperset. The paradigm also permits some hypersets to be practiced extensively over long periods of time.

Typical performance measures are: total number of trials needed to accumulate the criterion (e.g., ten) number of successful trials; and performance time, which is measured from home key release until depression of the final button for each trial, and then summed or averaged over the 10 successful trials. Unfortunately, the measures typical of the other paradigms reviewed here have not been systematically reported.

Hikosaka et al. (1995) defined three phases of learning in the $2 \times N$ task. The sequence discovery phase lasted until the first successful hyperset was generated. In the intermediate phase, successful trials intermixed with error trials. The advanced phase began when subjects showed virtually error-free performance. In addition to the decrease in errors, the speed of performance increased as a function of practice, but this improvement lagged the accuracy gain. The other notable result concerned retention. When well-practiced hypersets were retested after a one-month retention interval (without practice), performance time and accuracy were significantly better relative to new hypersets. Although the same was true after a six-month interval for performance time, error rates for old were not reliably lower than for new hypersets. This suggests a dissociation between order-controlling processes and speed-controlling processes.

Miyashita, Rand, Miyachi, & Hikosaka (1996) also described changing patterns of eye movements during the course of long-term $2 \times N$ task learning. After sufficient practice, anticipatory saccades (which started prior to target illumination and ended within the area of the next target) began to replace the visually guided saccades characteristic of early performance. Saccades preceded hand movements regardless of the stage of learning, but the anticipatory saccades occurred only for learned hypersets, and their frequency increased gradually as practice continued over 3–4 weeks. Video analysis revealed that hand movements preceding button-presses also became anticipatory – with the hand becoming poised over the next button awaiting its illumination. Movement time between first and second buttons in a set showed no decrease.

Each hand can be trained separately – providing a convenient method whereby laterality of control can be probed at various stages of the learning process. Rand, Hikosaka, Miyachi, Lu, & Miyashita (1998) investigated the effector-specificity characteristics of long-term sequence learning. After monkeys learned a particular hyperset with a selected hand, they were required to perform the hyperset with the

opposite (untrained) hand. Errors and performance time both increased modestly when the untrained hand was compared to the trained hand for the same hyperset. However, these increases were not to levels seen for new hypersets so these results indicate partial, but not total, transfer. This was not a laterality effect – the reversal was apparent for both left-to-right and right-to-left switches. Transfer was assessed only after very extensive practice of the hyperset. If the effect of learning becomes more effector-specific with greater amounts of practice, then greater transfer might have been seen at an earlier point in training. When the early learning period was later examined (Rand et al., 2000), the effector transfer results differed somewhat from those of well-learned hypersets. The second practice block of a new hyperset was performed with either the same hand used in the first block or the opposite hand. Error numbers did not differ between hands, but performance time was shorter for the practiced hand. Here again is a dissociation between serial order representation and speed. The latter appears somewhat effector-specific, even early in practice. The existence of partial transfer suggests that there remains some effector-non-specific memory that can be drawn upon to assist performance when required, consistent with results from other serial tasks, e.g., handwriting (Rijntjes et al., 1999; Wright, 1990). Recent research with keying tasks suggests that the development of an effector-specific learning component is critically dependent on extensive practice (Verwey & Clegg, in press; Verwey & Wright, 2004).

The first $2 \times N$ report using human subjects was an fMRI (functional magnetic resonance imaging) study by Hikosaka et al. (1996). The number of successfully completed sets as a function of imaging scans (i.e., practice blocks) was qualitatively similar to the pattern seen in the monkey data of Hikosaka et al. (1995). Presupplementary motor area (pre-SMA) was found to be particularly active during discovery of *new* sequences. This activation was unilateral, but the side differed between subjects. Moreover, the same location was activated when subjects repeated the experiment (thereby learning new hypersets). The pre-SMA was not active for individual movements in a control, pseudo-learning, condition. The supplementary motor area (SMA) proper showed only baseline activity during discovery, but became more active during performance of sequential movements.

Sakai et al. (1998) found that various cortical areas contribute to human sequence learning in the $2 \times N$ task. As learning proceeded, the relative levels of activation of these areas changed. There was a transition from strong activation in frontal areas to parietal areas. For instance, dorso-lateral prefrontal cortex (DLPFC, left and right) monotonically decreased in activation from early to intermediate to advanced stages of performance (as defined earlier), while activation in the IPS (intra-parietal sulcus) monotonically increased as performance became progressively better. Activation in the pre-supplementary motor area (pre-SMA) remained high through the first two stages, but decreased in the advanced stage. Another posterior (medial parietal) area, the precuneus, increased in activation from early to intermediate stages and then decreased again with further practice. A wave of heightened cortical engagement seems to pass from DLPFC to pre-SMA to medial parietal (precuneus) to lateral parietal (IPS) during the time course of sequence learning (cf., Koehlin, Ody, & Kouneiher, 2003).

Rand et al. (1998) explored the nature of the learned sequential representation. When monkeys had learned a hyperset, the individual sets were presented in reverse order. This manipulation greatly increased both error numbers and performance time (to almost new hyperset levels), indicating that the animals were not merely learning the order of button-presses for each set (visual configuration) individually, but were learning the transitions between sets and, perhaps, the whole hyperset as an extended sequence. This was also suggested by the prevalence of anticipatory eye movements made before next-set illumination (Miyashita et al., 1996).

Sakai, Kitaguchi, & Hikosaka (2003) reported on the spontaneous appearance of chunks of sets during human learning of 2×10 hypersets. In other paradigms – e.g., discrete sequence production, as reviewed below – chunk formation has been externally prompted by temporal grouping (or other pattern changes) within sequential stimulus presentation. The $2 \times N$ task has no such supra-set groupings, yet time gaps between performed chunks, each spanning 2 or more sets, emerged during practice of hypersets. Although overall hyperset performance time decreased as a function of practice, the clustering of sets into chunks became clearer and more consistent. The chunking patterns were independent of physical aspects of the sequence (like movement distance) and were different between subjects who learned the same sequences (cf., Verwey, 2003b; Verwey & Eikelboom, 2003). Subsequent to emergence of chunks, experimental rearrangements that respected chunk boundaries resulted in more accurate and quicker performance than rearrangements that violated them. The clear implication is that the chunks had become unitary elements (at one level of hierarchical control) that could be recombined fairly effectively as units, whereas attempts to break them apart and recombine them arbitrarily proved problematic. These results are similar to the transfer phase results of Verwey (1996). Probes of inter-hand transfer showed that the chunking patterns transferred from the non-dominant hand to the dominant hand, but not vice versa. The latter suggests an asymmetry in the intact brain's representation and storage of learned sequential chunks consistent with many prior results on language acquisition (e.g., Bullock, Liederman, & Todorovic, 1987).

Further work, focusing on the early learning period, demonstrated that order reversal had differential effects on error numbers and performance time (Rand et al., 2000). The second training block of a new hyperset was presented either in the original set order or in a reversed set order. For two of the four monkeys, second block set order had no effect on error number reduction. For the other two monkeys the error number reduction was greater for the same order condition. For all monkeys, there was a reduction in performance time for the same order condition but none when the set order was reversed. These and the earlier speed results suggest that a memory component that improves speed without reducing error rate is order-specific (as well as hand specific) and that this is the case throughout learning. Memory for order seems to differ between monkeys early in learning (and is not effector-specific), but this ceases to be the case later in learning, when memory for order becomes relatively effector-specific across all animals.

These memory dissociation results suggest that numerous substrates for learning and performance may exist within the brain. The powerful capability to directly

compare opposite extremes of the learning continuum within about 10 min has been employed to investigate the role of various brain sites in this form of serial learning using reversible focal inactivation techniques.

Miyachi, Hikosaka, Miyashita, Karadi, & Rand (1997) injected muscimol, a GABA agonist, to enhance local inhibition and thereby reversibly inactivate different zones of the striatum in the monkey basal ganglia. Inactivation of a striatal zone spanning the anterior caudate/putamen had a dramatic deleterious effect on discovery/learning of new sequences but no effect on performance of well-learned sequences. Injection into the middle-posterior putamen had a statistically detectable, but not overwhelming, effect on performance of well-learned sequences, but no effect on learning. The putamen may contribute to correct ordering of items within well-learned sequences, but other, extra-striatal, sites apparently assume most of the load in mediating recall and performance of well-learned sequences. This agrees with results of Lu, Hikosaka, & Miyachi (1998), who probed the role of the cerebellum in $2 \times N$ learning and performance. Muscimol was injected into zones of the dentate (most lateral) nucleus of monkey cerebellum. Unilateral injections into dorso-medial, dorso-lateral, and central dentate led to an increase in the number of errors on previously well-learned sequences when using the ipsilateral hand but not when using the hand contralateral to the injection. In contrast, the number of trials to criterion in the discovery/learning phase was not affected by muscimol injection into any dentate region for either hand. Injections into ventral dentate or other deep cerebellar nuclei (i.e., fastigial or interposed nuclei) likewise had no effect on trials to criterion. These results indicate zone-specific dentate inactivation effects on previously learned sequence production, but not on acquisition of new sequences. This pattern is similar to, but much stronger than, the effects observed by Miyachi et al. (1997) for middle-posterior putamen inactivation, and quite distinct from effects of inactivation of anterior caudate, which caused a marked retardation of new sequence acquisition.

Injections of muscimol at *all* tested deep nuclear sites led to elongated (intra-set) movement times for the ipsilateral hand. This is consistent with the widely held view of cerebellum as a predictive controller – absence of which entails a slow down in order to retain the requisite level of accuracy. For learned hypersets, anticipatory saccades decreased after muscimol injections into dorso-medial, dorso-lateral and central dentate as well as interpositus.

Nakamura, Sakai, & Hikosaka (1999) found that pre-SMA inactivation increased the number of errors for novel sequences, but not for learned sequences. SMA inactivation produced a similar, but insignificant, trend. Performance speed was hampered for both novel and learned sequences with both pre-SMA and SMA inactivation. However, the effect was larger for SMA inactivation.

Aggregation of results from the $2 \times N$ paradigm suggests participation of parallel substrates, such that the specific roles of different areas vary depending on the stage of learning. These results also imply organization above the set level: correct anticipations of the stimulus location to be pressed next in a not-yet illuminated set increased with learning.

2.1.4. Discrete sequence production (DSP)

Another fruitful paradigm is the discrete sequence production (DSP) task. In the DSP task, sequence length is limited to six or eight elements ('discrete') and key-specific cues usually follow a response immediately. DSP encourages internal control because of the limited sequence length and because there is usually only a limited number of sequences that can be distinguished and recognized easily. Therefore, DSP is better suited to study preparatory mechanisms, hierarchical control, and sequence segmentation than tasks in which sequences are longer and have no clear beginning, such as the SRT task. Also, the effect of serial position can be studied in DSP. In DSP tasks, it seems fair to assume that participants use preparation of at least the first few elements of forthcoming sequences to improve performance.

As already noted, many psychologists have long suspected that the control of motor behavior is hierarchical (e.g., Miller et al., 1960). Restle (1970) and Povel & Collard (1982) inferred control by a hierarchical representation because more errors occurred at some than at other positions in the sequence. Using a particular version of a DSP task, Rosenbaum and colleagues performed a series of studies in which participants chose between alternative sequences of keypressing responses (see Rosenbaum, 1987 for a comprehensive overview). In these *choice reaction time* (cRT) tasks, Rosenbaum, Saltzman, & Kingman (1984) asked participants to choose between "i" vs. "I", "ir" vs. "IR", or "irm" vs. "IRM", where "i", "r", and "m" denote key presses of the index, ring, and middle fingers of the left hand, respectively, and "I", "R", and "M" denote key presses of the right index, ring, and middle fingers, respectively. Participants learned to associate one visual signal (O) with one sequence and another signal (X) with the other sequence but were not highly practiced with any sequence. On each trial, one of the two signals appeared and the participant was supposed to produce the designated sequence as quickly as possible. The sequences in the experiments were chosen because they are easily organized in a hierarchical fashion. If any sequences would be controlled hierarchically, these would be.

The timing results in such tasks suggest hierarchical control (also see Gordon & Meyer, 1987; Kornbrot, 1989). The main results in Rosenbaum, Saltzman, et al. (1984) were that latency T_1 increased with the number of keypresses in the sequence, and that the mean time for the second key press T_2 was longer when that response was embedded in a sequence of three than when embedded in a sequence of two. These results extend the SLEL and SLER, observed earlier for sRT tasks (Sternberg et al., 1978), to cRT tasks. A subsequent experiment showed that the latency of a sequence was heavily influenced by the alternative sequence. For instance, "I" was considerably longer when the alternative sequence was "irm" than when it was "i". It appeared also that selection time takes longer as sequences differ in more features (e.g., direction, hand). Apparently, just-used motor programs are preserved so that the features distinguishing the just-used program from the next program to be performed can be changed, and it takes more time to change these features than to leave them unchanged (Rosenbaum & Saltzman, 1984). By carefully manipulating properties of alternative sequences, like the position of sequence differences (e.g., "irm" vs. "iRm" and "irm" vs. "irM"), Rosenbaum, Inhoff, & Gordon (1984) showed that T_1 increases more as (a) there are more decisions (which of two alterna-

tive responses is to be used) in the sequence, (b) these decisions are earlier in the sequence, and (c) the rules used to select alternatives are more complex. For example, T_1 was shorter when alternative sequences were mirror images (e.g., “iim” vs. “IIM”) than when they were not (e.g., “iim” vs. “IMM”). These data suggested that participants prepare a plan and that choices are made at the highest hierarchical level possible (i.e., affecting as many elements in one step as possible).

These and several other results were encompassed by the Hierarchical Editor (HED) model (Rosenbaum, Inhoff, et al., 1984), which proposes that participants prepare for a choice between response sequences by establishing an abstract program with all the features common to the possible sequences. The program is hierarchically organized in that common features are represented by a single node in the hierarchy whereas uncertainties are indicated by a subtree that represents the alternative features. Before the reaction signal appears the hierarchy is traversed from the top node, via the common feature nodes, to the subtree that represents the first uncertainty. After the reaction signal is identified but before the first element is executed (i.e., during T_1) all remaining nodes are traversed and uncertainties are resolved. Response elements are not executed yet. This phase prepares the hierarchy for execution and is called the *edit pass*. Next, control returns to the top of the hierarchy and the traversal process begins anew, this time executing each sequence element encountered. This is the *execution pass*. Given the assumption that traversal of one to the next node (i.e., a processing step) takes a finite amount of time, predictions can be made on the relative times between successive keypresses. However, position-dependent IRI effects seen in Sternberg tasks are not explained (cf., Boardman, 1995; Boardman & Bullock, 1991).

Another version of the DSP task has been studied by Verwey and colleagues. In this research, emphasis is on the processing mechanisms that become engaged as participants extensively practice keypressing sequences of up to six elements. In this paradigm, participants practice the sequences in a choice RT (cRT) task in which each of the alternative sequences is practiced for about 500 trials. In contrast to those used in the Rosenbaum studies, these sequences do not involve any apparent regularities that can be expected to induce hierarchical representations. Given that the interest is in execution rate as a function of serial position, confounding of finger-specific effects with serial position is prevented by rotating fingers at each serial position across participants. So, across all participants each serial position gets the same contribution from each of the fingers used in the experiment. To help participants learn these sequences, each keypress is typically indicated by a cue that immediately follows depression of the preceding key. So, whereas participants start off responding to apparently unrelated cues they gradually learn to produce the keypresses as a sequence in response to the first cue and ignore later cues (Verwey, 1999).

Initial studies in this research line used highly practiced keying sequences to assess the earlier notion that extensive practice yields an integrated representation for the entire movement sequence, that is, a motor chunk that can be selected and executed as a whole (e.g., van Mier & Hulstijn, 1993). A first study involved a two-choice sequence production task (Verwey, 1994). One sequence contained four, the other two keypresses. The main results were (a) reduction with practice of the difference

between the times to initiate long and short sequences (i.e., reduction of the SLEL), (b) a larger practice effect on final keypresses than on earlier keypresses, and (c) a larger practice effect on the final keypress of the long than of the short sequence. The results were in line with the notion that practice allows the processes that are required for executing individual elements to overlap with execution of the preceding element. Evidence for one hypothesized type of motor chunk development, an increasing execution rate with serial position due to an increasing spread of activation across sequence elements (MacKay, 1982), was not found (also see Verwey, 1999; Verwey & Eikelboom, 2003).

The absence of an increasing execution rate with position does not imply an absence of motor chunks, conceived broadly as integrated sequence representations. Verwey (1996) tested this by having participants produce a series of nine keypresses in rapid succession. This task has some resemblance with SRT tasks (see next section) but it involved more practice and it replaced the 200ms RSI typical of SRT with zero RSIs after most responses. However, RSIs of at least 500ms were used at two or three fixed positions of the 9 possible. These few RSIs were assumed to determine the boundaries of any motor chunks that might develop. The development of motor chunks qua sequence-specific representations was first tested by requiring participants to execute the entire 9 item sequence as rapidly as possible in a transfer phase. Despite this instruction, they exhibited long pauses at the locations where they had previously been exposed to the long RSI (see Fig. 3). Moreover, performance rate slowed when the pauses were located at non-practiced positions. These and other data were interpreted as strong evidence for motor chunk development. Another important finding was that the use of motor chunks may be concealed when there is time for advance preparation and chunks involve no more than about four keypresses.

In line with earlier findings by Brown & Carr (1989), the Verwey (1994) study showed that the fourth element in the 6-key sequence was relatively slow in early practice, and that this disappeared with more extensive practice. One interpretation is that longer sequences consist of independent segments and that practice allows them to be more efficiently concatenated, presumably by enabling preparation of one segment during execution of the previous segment. If so, disappearance of the relatively long interval halfway through the sequence is attributable to increasing overlap between preparation of one and execution of the preceding segment. To examine this hypothesis, Verwey (1995) investigated whether at least one particular process, response selection, could overlap with, and complete during, the execution of familiar keying sequences. To that end, subjects pressed a *familiar* series of keys prior to pressing one *stimulus-dependent* key. Because it is known that the effect of S–R compatibility does not disappear with practice, response selection demands were manipulated by utilizing spatially compatible or incompatible stimulus–response mappings. After practice, the time increment initially needed to select a key using an incompatible mapping vanished from the timing data when the number of keys in the sequence that preceded the choice key was four rather than two. This finding indicates that response selection operations associated with incompatible S–R mappings can occur concurrently with execution of a sufficiently long sequence

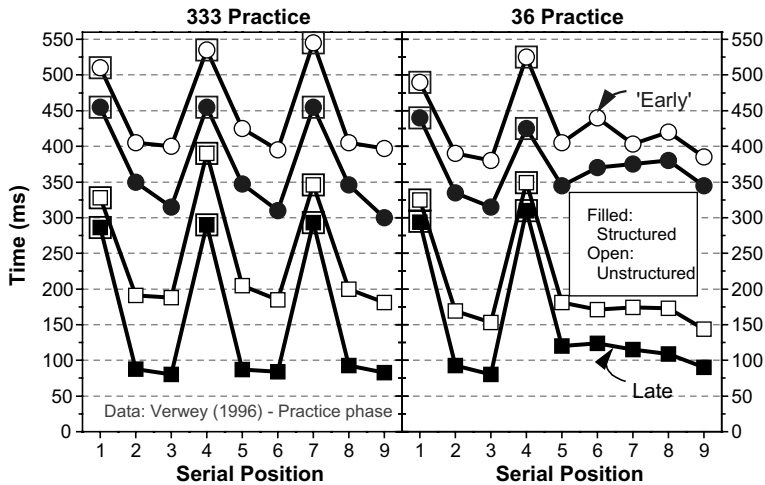


Fig. 3. Practice phase results from Verwey (1996, fig. 2). The panels plot latencies (akin to reaction times) and inter-response intervals (IRIs) as a function of serial position and of imposition (Structured) or not (Unstructured) of temporal delays during sequence performance. Left: Results from the temporal delay pattern (333 condition) that produced three 3-item response groups. Right: Results from the temporal delay pattern (36 condition) that produced a 3- and a 6-item response group. Larger bounding squares indicate sequence start (SS) times. Circles and dark grey lines indicate moderate levels of practice (despite Verwey's 'Early' label); squares and black lines represent advanced ('late') levels of practice. Over the course of practice an obvious increase in overall performance speed is apparent, as too is loss of SLEL and IRI inhomogeneity (when averaged across subjects), while SLER survives. The ratio effect is present in all these plots based on practiced performance but is not present when subjects are producing novel sequences in the DSP paradigm.

and thus can become invisible in timing data. Similarly, the transition between independent segments of long sequences may become invisible in timing data over the course of practice.

Verwey (2001) investigated this issue more directly by having participants first practice several keying sequences, and then produce two of these sequences in rapid succession. In contrast to Verwey (1995), when a single keypress quickly followed a familiar keying sequence, the transition between the familiar sequences remained relatively slow. By manipulating stimulus-sequence mappings, it was demonstrated that the selection of the second sequence still occurred during execution of the first sequence. Therefore, the relatively slow transition between the first and second sequence suggests that processes that follow chunk selection, and that are not used when a single response is selected, cannot overlap with execution of the preceding motor chunk. The results also demonstrated that once formed, motor chunks are robust. When participants performed two 2-key sequences repeatedly in rapid succession, they maintained their use of distinct 2-key segments, although they could have prepared and carried it out as a single 4-key sequence (cf., Sternberg et al., 1978).

As discussed earlier, Rosenbaum, Saltzman, et al. (1984) had shown that sequence execution may be influenced by the alternative sequence. This led Verwey, Lammens,

& van Honk (2002) to examine the hypothesis that when one of two 6-key sequences consists of a repetition of a 3-key segment, the other 6-key sequence might be segmented also as two 3-key parts. The results supported this hypothesis in that the interval between the third and fourth keypress in both sequences was relatively long. This gave rise to the hypothesis that perhaps each participant segments longer sequences, but that this is concealed because of individual segmentation differences when the alternative sequence does not consist of a repeated 3-key sequence. A preliminary indication for this suspicion is the well-known finding that the mean element execution time increases with sequence length, i.e., the SLER (Sternberg et al., 1978). This hypothesis was investigated in two studies. First, Verwey & Eikelboom (2003) had participants practice a 3- and a 6-key sequence. Analyses across serial position and participants confirmed the SLER for the 3 vs. 6 comparison. Further analyses supported segmentation for 6-key sequences with and without clear regularities at various levels of practice and irrespective of a preceding task. This suggests that a component of the SLER is caused by segmentation of longer sequences and that motor chunks represent sequences of a limited length only. A second study examined the SLER across sequences of lengths two, four and six keypresses (Verwey, 2003b). Detailed analyses confirmed the upward curvature noted first by Sternberg et al. (1978), and showed that in each 6-key sequence the rate effect was caused by a few slow elements while the fastest elements in these 6-key sequences were no slower than corresponding (always non-initial) items in shorter sequences. A concurrent memory task did not affect the transition between segments, suggesting that controlling the transition does not require working memory and perhaps depends on a higher order sequence representation that is able to code longer sequences. If so then even performance of longer sequences that do not involve regularities may engage hierarchical control.

These findings were taken as support for a dual-processor model. In it, a motor processor executes representations that we may still call motor chunks, which control highly practiced segments of limited length. A cognitive processor either triggers each element in parallel to the motor processor, or concatenates the segments in the sequence (Verwey, 2001). Given the high execution rates that are gradually attained with practice, these motor chunks seem to involve a type of coding that requires little further processing to execute the sequence. A neural interpretation of this model has been proposed on the basis of the finding that transcranial magnetic stimulation (TMS) of the premotor cortex slowed intervals at the start of, and within, a segment (Verwey et al., 2002). It assumes that motor chunk length is limited by the number of keypresses the supplementary motor area/motor cortex loop can handle. With longer sequences the basal ganglia would have to concatenate these cortically stored segments via a relatively slow thalamo-cortical motor loop, thus inducing hierarchical control of longer keying sequences, or triggering of individual elements. This mention of subcortical structures raises an interesting parallel with the SRT task (see below). Even though participants in this version of the DSP task are fully aware that they are executing a few fixed sequences, and despite the extensive amount of practice, there are always a few participants who at the end of the experiment are not able to give a full account of the sequences they just executed, even

though their performance level is not below that of aware participants (Verwey, 2003a, 2003b). This finding suggests the hypothesis that the sequence elements that allow fast sequence production in the DSP are not accessible for translation in other codes either, just as in the SRT task.

More recent findings suggest more than two control levels. First, Verwey & Wright (2004) showed that, in addition to a verbal and a more automated control level, extensive practice yields a third component that is effector-specific in that it does not transfer when other fingers are being used (cf., Rand et al., 1998). This finding might indicate parallel control by a third processor, though it might also be caused by the development of a timing pattern that is optimally adjusted to the effectors used. Second, indications for more than two processing modes are also provided by a study in which participants unexpectedly were required to produce familiar sequences (Verwey, 2003a). Detailed analysis of the individual interkey interval (IRI of Sternberg et al., 1978) distributions showed evidence, for most participants, of three separate peaks in the distributions. These data suggest that familiar sequences can be carried out in three different modes among which the participants switched once the familiar sequence was recognized. Third, detailed analysis showed that a memory task slowed sequence initiation and execution of the second response in a 2-key sequence, but not in longer sequences and not the transition between segments (Verwey, 2003a). Given that longer sequences are segmented and, hence, already involve two control levels, this finding might indicate a third control level in which working memory is involved. Even though more evidence for a three-level control model is required, a three-level control model fits the notion that control of individual segments at the cortical level, and concatenation of these segments by the basal ganglia, are extended by executive control provided by the frontal lobe.

At first sight, the Rosenbaum et al. and the Verwey et al. studies agree on the use of hierarchical control. However, a distinction should be made between two types of hierarchical control (Broadbent, 1977). Hierarchical control according to Rosenbaum, Inhoff, et al. (1984) involves a single processor traversing a hierarchical representation. This may well be more suited for modeling relatively unfamiliar sequences. In contrast, hierarchical control implicated by the Verwey et al. results involves different processors working in parallel at different levels of sequence descriptions. In fact, this type of model involves two further subtypes of models. On the one hand, low level segments are carried out by a dedicated motor processor while a higher level processor is preparing forthcoming segments, thus being responsible for concatenation of the segments. On the other hand, various processors may race to trigger each forthcoming element. Depending on the level of practice, some processors maybe more successful, and hence may allow certain processors using particular codes (like spatial, or motor) to be dominant. Learning of multiple representations is in line with theoretical perspectives on highly practiced tasks like writing, speech, typing, and key tapping (e.g., Hulstijn & van Galen, 1988; Klapp, 1995; MacKay, 1982; Newell, 1978; Semjen & Garcia-Colera, 1986). All in all, there is ample reason to believe that with practice sequence representations develop at various levels, probably involving different task dimensions (Keele, Ivry, Hazeltine, Mayr, & Heuer, 2003) and that the production of movement sequences can be flexibly

adjusted to the availability of representations and processing resources like attention and working memory.

2.1.5. Serial reaction time (SRT)

The development of the SRT task started in 1987 (but see Bahrnick, Noble, & Fitts, 1954 for an early version) when Nissen & Bullemer (1987) wondered whether non-episodic forms of memory would perhaps be less dependent on attentional processing than episodic memory forms. They put this to the test with a task in which a light appeared at one of four locations on a video monitor. As quickly as possible, participants pressed the key, out of a set of four, that was directly below the position of the light. Each time a response was given, the next stimulus was presented after a response to stimulus interval (RSI), typically of about 200 ms. A particular 10-trial sequence of light positions was practiced for 10 blocks. Each block consisted of 10 sequence cycles so that naïve participants would not detect the transition between the end of one and the start of the next repetition. Nissen and Bullemer established a divergence between RTs in the sequence vs. the random order condition after only six repetitions of the sequence, and a 50% RT reduction in the sequence over the course of 40 min of practice whereas hardly any improvement was found in performance of the random sequence. However, when low and high pitched tones were presented before each of the keypresses and participants were instructed to count the low tones, RTs declined with practice to a similar extent for a sequence and a random group. Participants appeared to learn to combine the two tasks, but they did not learn the sequence. This was confirmed by a subsequent test of episodic memory, known as the generate task, in which participants were asked to press the key corresponding to where the next stimulus would appear in the sequence. Nissen & Bullemer (1987) concluded that attention is required for learning sequences regardless of whether learning involved episodic or non-episodic memory. The results were considered in line with the use of distinct memory systems because verbalizability of the sequence was severely hampered in amnesics with Korsakoff syndrome even though performance was not. Later studies demonstrated sequence learning also, in that RTs increased dramatically when the repeating pattern was modified (e.g., Cohen, Ivry, & Keele, 1990).

Many SRT participants do not express awareness of the repeating sequence. Participants with awareness can express their sequence knowledge in various ways and are said to have *explicit knowledge*, whereas participants who cannot express that knowledge in other ways than by performing the task are said to have *implicit knowledge*. This distinction led to the question of whether participants showing awareness of the sequence use another memory system than participants showing no signs of awareness. The idea of different memory systems has been rejected by some researchers because awareness (i.e., presence of explicit knowledge) was inadequately assessed in early studies. Typical methods were (a) free verbal sequence reproduction, (b) prediction of the next elements (using the so-called generate task), and (c) recognition of short stimulus segments from a set of which only some had been practiced. Shanks & St. John (1994) argued that free verbal reports do not constitute sufficiently sensitive tests of explicit knowledge because people may not be

able to express weak explicit knowledge in such a task. They argued that forced-choice tests, such as sequence generation, and segment/sequence recognition, should be used to assess whether participants have explicit knowledge (also see Jimenez, Mendez, & Cleeremans, 1996a, 1996b; Perruchet & Amorim, 1992). Because studies have shown strong associations between forced-choice and SRT performance levels, it was argued that the dissociation between SRT performance and verbal reports is caused by a methodological flaw and, hence, that the evidence for independent memory systems was unjustified. However, by now, there are various indications that implicit and explicit knowledge are based on functionally different forms of memory.

For example, Destrebecqz & Cleeremans (2001) argued that no task, not even a forced-choice task, can be assumed to be process-pure, in the sense that it involves solely implicit or explicit knowledge. Findings that aware participants are better at predicting oncoming cues in the generation task, and better at recognizing sequence segments, cannot be exclusively attributed to the influence of explicit knowledge because implicit knowledge can also affect performance on these tasks. In order to circumvent this problem, the process dissociation procedure (PDP; Jacoby, 1991) was adapted for the SRT task (Destrebecqz & Cleeremans, 2001). In the PDP, participants are informed after practice that there had been a repeating sequential pattern in the input, and they are asked to freely generate a series of 96 trials that 'resemble the training sequence as much as possible' (an *inclusion* condition). Participants are instructed to rely on intuition if necessary. Then they generate another 96 trials, but this time they try to *avoid* reproducing the sequential regularities of the training sequence (an *exclusion* condition). The results of the PDP confirmed the earlier mentioned associations between performance levels on forced-choice tests and the SRT task. However, analyses also revealed that participants who practiced with zero RSIs (assumed to suppress development of explicit knowledge), produced significantly more fragments of the training sequence in the exclusion condition than participants from the 200 ms RSI condition. Participants' inability to exclude familiar sequences in the exclusion condition suggests that performance in the inclusion condition was in part based on implicit sequence knowledge, and that this was stronger when practice involved a 0 ms RSI. The results were seen as support for a functional dissociation between implicit and explicit learning although perhaps not for entirely different memory systems.

A second study supported the hypothesis that aware participants are faster than unaware participants because they use sequence knowledge to prepare forthcoming responses before the cue is presented. Earlier research had already suggested this (e.g., Willingham, Nissen, & Bullemer, 1989), but Eimer, Goschke, Schlaghecken, & Stormer (1996) found support using electroencephalography (EEG). At the half-way and end points of a 28-block SRT task, they asked participants whether they had noted anything special and, if they referred to regularities, asked them to reproduce those regularities. At the end of training they were asked also to identify the correct sequence from a set of six alternatives in a forced-choice test. On the basis of these subjective results, participants were classified as unaware or having some awareness. In the first experiment, for instance, seven participants qualified as partially aware or aware (i.e., they were eventually able to identify at least four

successive items of the sequence), and nine as unaware. Event-related potentials (ERPs) were extracted from the EEG recorded during performance of a SRT task that included a deviant stimulus on half (Experiment 1) or all of the sequences (Experiment 2). Because the so-called N2 component of the EEG is known to be enlarged with unexpected events, this allowed a non-intrusive on-line measure for the development of awareness (expectancies). Participants who showed explicit knowledge on the basis of the forced-choice sequence recognition task developed larger deviance-related RT effects and also produced an enlarged N2 effect, relative to participants lacking explicit knowledge. Furthermore, the N2 effect increased with practice only for the more aware participants. It was concluded that the deviance-related N2 indicates gradual development of explicit knowledge. In line with the notion that aware participants base their higher execution rate on preparation of the individual responses during RSIs, the lateralized readiness potential (LRP) in the EEG (an index for response preparation) showed evidence for early response activation in aware participants but not in unaware participants. These results confirm that awareness is associated with higher execution rates because explicit knowledge is used for preparing forthcoming responses. Such preparation exemplifies one way that aware participants are more flexible in using sequence knowledge than unaware participants. Overall, the study validates awareness classifications based on behavioral (forced choice) tests with an independent, physiological measure.

Other neuropsychological studies provide evidence for the notion that different brain areas are involved in explicit and implicit knowledge. The original Nissen and Bullemer finding that verbalizability of the sequence was severely hampered in amnesics with Korsakoff syndrome while performance was not, was replicated for several neurological diseases (Alzheimer's disease, Down and Korsakoff syndrome; Ferraro, Balota, & Connor, 1993; Vicari, Bellucci, & Carlesimo, 2000), while the reverse pattern (poor SRT performance but good verbalization) was associated with other disease syndromes (Parkinson's and Huntington's disease, lesions of the basal ganglia; Doyon et al., 1997; Vakil, Kahan, Huberman, & Osimani, 2000). Together these studies imply a double dissociation between tests of implicit and explicit knowledge across studies. Consistent evidence comes from a single study in which amnesic patients exhibited superior implicit sequence knowledge (though after more practice) than healthy control participants, but less explicit knowledge as indicated by a forced choice between the practiced and five alternative sequences (Reber & Squire, 1998). The authors proposed that explicit sequence knowledge is supported by medial temporal lobe structures that mediate declarative knowledge, whereas implicit sequence knowledge is supported by various other brain areas (e.g., the neostriatum, supplementary motor area, and motor cortex).

Many studies have used brain scanning methods to determine whether awareness in the SRT task is related to activity in particular brain systems. Neuroimaging with positron emission tomography (PET) during SRT task performance indicated that the sensorimotor cortex and neostriatum were active in conditions in which little explicit knowledge was acquired (Grafton, Hazeltine, & Ivry, 1995; Hazeltine, Grafton, & Ivry, 1997; see Cleeremans, Destrebecqz, & Boyer, 1998 for an overview), whereas other structures were indicated when the sequence was learned explicitly. To further

reveal structures involved in explicit learning, Destrebecqz et al. (2003) combined PET and the earlier described process dissociation procedure (Destrebecqz & Cleere-mans, 2001). The results showed that activity in the anterior cingulate/medial pre-frontal cortex was exclusively correlated with the explicit component of performance during recollection of the learned sequence. However, it is not clear whether differential brain activity necessarily implies that different memory systems are involved in implicit and explicit sequence knowledge.

Keele et al. (2003) reviewed an abundance of empirical evidence, much of it from SRT tasks, to motivate a dual-substrate theory of sequential representation. One dorso-medial system includes parietal and supplementary motor areas, whereas a second, ventro-lateral system includes temporal cortex and lateral prefrontal cortex. Thus of the four cortical areas mentioned, two, the parietal and temporal, are late stages of the well-known dorsal and ventral (sensory) processing streams (Goodale & Milner, 1992; Ungerleider & Mishkin, 1982). Each proposed system learns sequential regularities by association. Dorso-medial learning is proposed to be implicit whereas ventro-lateral learning is proposed to be explicit *and* implicit. The dorso-medial system consists of encapsulated modules each of which extracts sequential regularities in a single dimension to which it is attuned (cf., Adi-Japha & Freeman, 2000). Attention is not needed for this automatic extraction, and the uni-dimensional modules are immune to potentially disruptive information in other dimensions. Although the authors offer no formal definition of 'dimension', dimensions are proposed to span attributes of stimuli within a particular modality (cf., Treisman, 1988), and attributes of the motor system (e.g., hand vs. feet). Ventro-lateral learning, whether implicit or explicit, builds associations between multidimensional events. Such learning facilitates acquisition of complex sequences, but makes sequence learning vulnerable to overload by uncorrelated dimensions. To protect from such overload, only signals specified as relevant by the current task set – i.e., the attended signals – can enter the multidimensional system and participate in learning. Secondary tasks are predicted to disrupt multidimensional learning not by inducing capacity limitations, but by disrupting coherence between successive events. However, once events are admitted by the attentional system, associative learning is implicit: it will automatically operate on all admitted signals. Learning may also become explicit because attending to events implies that they are accessible to processes underlying awareness.

The primary difference between aware and unaware individuals seems to lie in the capacity of aware individuals to flexibly adjust sequence knowledge. They can prepare forthcoming responses in a familiar sequence, recognize short segments, and produce verbal sequence descriptions. This capacity may depend on having a type of representation – an explicit representation – that unaware participants lack. Explicit sequence representations are flexible, rapidly forgotten and resistant to sleep deprivation. Unaware participants have only implicit representations, which are task and context specific, robust to forgetting and susceptible to sleep deprivation (Heuer, Spijkers, Kieswetter, & Schmidtke, 1998). With extensive practice they may even comprise an effector-dependent component (Verwey & Clegg, *in press*). That these forms of representations are functionally different does not necessarily mean that

they involve different memory systems. However the differential associations with cortical brain areas, combined with the overwhelming evidence that all areas of the cerebral cortex possess local memory, strongly suggests that there are multiple memory systems subserving implicit and explicit knowledge of sequences.

2.2. Modeling approaches and the history of competitive queuing

From at least the time of [Lashley \(1951\)](#), cognitive scientists have marshaled evidence in support of the thesis that fundamentally parallel representations underlie much of our learned serial behavior. Such behavioral evidence provided a basis for the proposal ([Grossberg, 1978a, 1978b](#)) of a class of parallel sequence production models that have since come to be known as *competitive queuing* (CQ) models ([Bullcock & Rhodes, 2003](#); [Houghton, 1990](#)). Such models (see [Fig. 4](#)) follow naturally from two assumptions: (1) more than one plan representation can be simultaneously active in a planning layer; and (2) the most-active plan representation is chosen, in a second neural layer, by a competition run to decide which plan to enact next. In CQ models, activation is the ‘common currency’ used to compare alternative plans, and simple maximum-finding or WTA (winner-take-all) dynamics can be used as the choice mechanism in the choice layer. Once a plan wins the competition and is used to initiate a response, its representation is deleted from the field of competitors in the planning layer, and the competition is re-run. This iteration allows the two-layer network to transform an initial activity distribution across plan representations, often called a *primacy gradient* ([Grossberg, 1978a, 1978b](#); [Page & Norris, 1998](#)), into a serial performance without any reliance on associative links among elements of the sequence. More generally, and especially in applications to linguistic sequences, it is recognized that the rank ordering of plans is subject to dynamic reordering during production. Thus, it is the emergent gradient, rather than an initial primacy gradient, that controls overt sequence production ([Houghton, 1990](#); [Ward, 1994](#)). For simplicity, the following discussion will refer to CQ applications that do not utilize dynamic reordering (except where [Ward’s](#) work is introduced). It should be noted that the standard CQ process of deleting already-executed plans to ensure progression through the entire sequence is not considered a form of dynamic reordering.

The primacy gradient across plan representations in a CQ model is a fundamentally parallel representation of serial order. Thus, CQ models provide a much different basis for control of serial behavior than so-called recurrent neural networks (RNNs), one modern descendent of associative chaining models. An RNN, in this usage, is a network in which each output is fed back as one component of a high-dimensional input (or other pre-output) stage. After extensive sequence-specific learning, this fed back (recurrent) signal combines with other state information to create a distinctive context for eliciting the correct next output. An RNN’s representation of a learned sequence is therefore fundamentally serial, in the sense that the information that specifies the sequence only becomes available as the serial performance unfolds. In contrast, all the information needed to specify a forthcoming sequence is present in the current state of the planning level of a CQ system and no learning is needed to represent the sequence: the primacy gradient suffices. Having

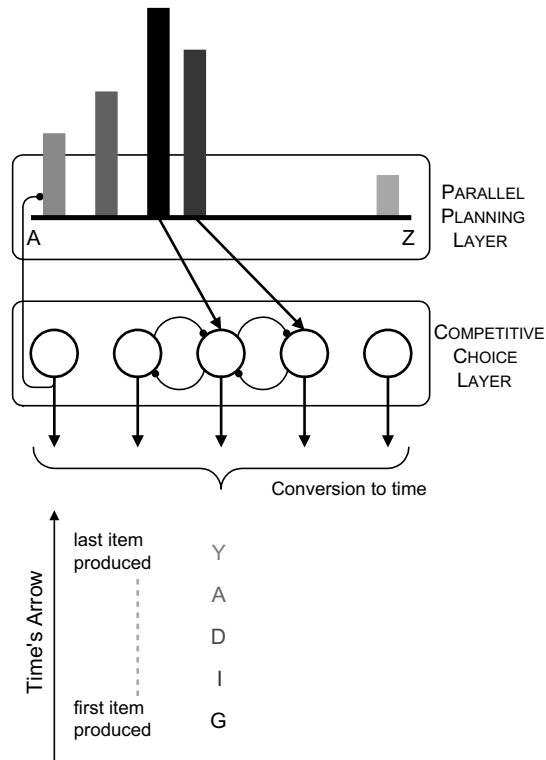


Fig. 4. Initial state of a two-layer competitive queuing (CQ) system, prior to production of a five letter sequence. The sequence that will emerge is shown in the lower part of the figure. Excitatory connections terminate with arrowheads, inhibitory connections with filled circles. The most-active plan is selected for execution in the lower, competitive choice, layer by a winner-take-all dynamic whose outcome is wholly determined (in the absence of noise) by the activation gradient (representing the to-be-performed sequence) present in the parallel planning layer. Once a plan representation wins at the competitive layer, a large output signal is sent to initiate execution of the corresponding response (descending arrow) and to delete the plan's representation in the parallel activation layer (ascending path to parallel planning layer). This process iterates until all plans have been enacted and all planning layer activities deleted. The result is sequential plan execution that corresponds to the initial rank ordering (gradient) of plan activation levels in the upper field of the CQ network. Although each competitive layer node would send an inhibitory connection to its correspondent in the parallel planning layer, only one such connection is shown here, to avoid clutter. In this example, which uses recurrent inhibition in the choice layer, each competitive layer node would inhibit all others, but only nearest-neighbor inhibition for a single node is actually depicted. From Bullock and Rhodes (2003).

such an explicit, parallel, activation-based representation of (even novel) sequential plans is advantageous for many purposes. For example, such representations can be learned and recalled via the compressive and expansive coding operations that may underlie the phenomenon of motor chunking.

Recent neuronal recordings in frontal cortex (e.g., Averbek, Chafee, Crowe, & Georgopoulos, 2002; Cisek & Kalaska, 2002) have strikingly confirmed *four key predictions* of the CQ class of models as originally proposed in Grossberg (1978a,

1978b). Notably, the study of Averbeck et al. (2002), Averbeck, Chafee, Crowe, & Georgopoulos (2003), and Averbeck, Crowe, Chafee, & Georgopoulos (2003) showed (1) that prior to initiating a serial act (of using a cursor to draw a geometric form with a prescribed stroke sequence – prescribed sequence figure drawing, PSFD), there exists an active parallel (simultaneous) representation of all the strokes planned as components of the forthcoming sequence. Also, (2) the strength of activation of a stroke representation predicts its order of production, and (3), as the sequence is produced, the representations are serially deleted at the times that the corresponding strokes are enacted. Several studies (Averbeck et al., 2002; Basso & Wurtz, 1998; Cisek & Kalaska, 2002; Pellizzer & Hedges, 2003) also give evidence for (4) partial activity normalization. The amount of activation that is spread among the plans grows more slowly than the number of plans (in the sequence), and eventually stops growing. This was hypothesized (Grossberg, 1978a, 1978b) to result from competitive interactions among simultaneously active plans, and it places a low upper bound (e.g., five plus or minus two) on the number of plans that can be simultaneously active in a motor working memory for sequences. This upper bound – perhaps as low as four on average – is a property of human working memory as assessed in immediate serial recall (ISR) paradigms (Cowan, 2000), though in CQ models in this area (e.g., Burgess & Hitch, 1992; Page & Norris, 1998), the partial normalization is more often conceived of as resulting from a process of time-based decay, consistent with Baddeley & Hitch (1974).

Simulations of CQ models (Boardman & Bullock, 1991; Rhodes & Bullock, 2002b) have shown that they can readily explain the SLEL (Sternberg et al., 1978), as well as the characteristic pattern of response times – the ratio effect – that is taken as a chronometric signature of the *collective planning strategy* (Conway & Christiansen, 2001). To recall, in the Sternberg et al. (1978) task, subjects were told to repeat short novel prepared lists as fast as possible following an external signal. This qualified it as a working-memory dependent sRT task. A related list-recall task is the ISR task (as described earlier), in which subjects also recall a short novel list from working memory, but without explicit instruction to initiate or perform recall as fast as possible. This non-RT sequence production task has also been modeled successfully within the CQ framework. In particular, to the CQ assumptions noted above (primacy gradient, deletion upon enactment, and iterated competitive choice of most-active remaining plan), Page & Norris (1998) added two further assumptions: that the choice is noisy, and that decay of activity in the planning layer occurs during input to the planning layer and during intervals spent performing items from the list. Error data favor both assumptions, and this extended model was able to address data on errors of serial recall. One kind of error, simple failure to recall, is most probable for list-final plans in long sequences. The extended model explains this as a consequence of their low initial activation level (due to being last in the primacy-gradient-coded sequence), which in turn makes them more susceptible to falling into inactivity due to the decay that can occur during enactment of the prepared sequence. Another feature of error data from ISR studies is that the majority of transposition errors (items are recalled, but in incorrect order) are simple exchanges with immediately adjacent items in the planned sequence. Given moderately noisy choice,

this likewise follows from the gradient representation, because noise in the choice layer is less likely to illicitly promote a plan by two positions in the activity gradient than by one position. Moreover, whenever a transposition is an adjacent item exchange, then the earlier occurring of the two items is followed not by the correct next item in the target sequence but by the prior item from the target sequence. As noted earlier, such ‘fill in’ (Page & Norris, 1998) by the prior item is predicted by the CQ model, but is the opposite of what is expected by associative chaining models, either in the classical or RNN incarnations. Chaining predicts that any item that appears too early should be followed by the next item in the target sequence, not by the skipped item. But filling in by the skipped item is much more likely in the ISR task.

When Lashley (1951) argued against the associative chaining theory of serial order, he implicitly assumed that the elements of sequences are not coded with respect to their context. For example, in a context free code, the “u” in “struck” is represented in the same way as the “u” in “crust”. Since associative chaining models operate by following explicit links between the elements of sequences, the non-contextual “u” cannot be followed by both the “ck” in “struck” and the “st” in “crust”. Wickelgren (1969), however, pointed out that if the elements of sequences are coded in a context specific manner, associative chaining models do not have this difficulty. For example, if the “u” in “struck” is coded as “_ru_{ck}”, and the “u” in “crust” as “_ru_s”, the ambiguity is resolved. Thus, if associative chaining works with a context sensitive code, it is able to overcome some of the difficulties raised by Lashley (1951). The context sensitive code clearly predicts that a sequence element will be coded differently, depending upon the sequence in which it is being executed. Interestingly, as we review below, the neurophysiological representation of sequence elements is sensitive to the sequence being executed. Specifically, neural responses for a specific sequence element differ, depending upon both the position of that element in the sequence, and the sequence elements which precede and follow it. These data do not favor either the associative chaining or CQ hypotheses, but they do suggest that the brain is using a context sensitive code.

It might be thought that the CQ model cannot apply to syntactic language production, because sequencing errors in language production often violate the ‘adjacent items exchange’ error pattern that predominates in ISR studies (which typically use non-grammatical item sequences). In most sequencing errors in language production, exchanges respect grammatical constraints, as when a sequencing error transforms the intended “flying saucers” into the spoonerism “sighing flosers”. Note that the same example supports the CQ postulate that the initial segments of both words were already co-active in a planning field prior to production of either word. Moreover, it is plausible that the exchange error occurred because noise transiently rendered the plan for “fl” less active than the plan for “s” at the instant that “flying” should have been spoken. In fact, several neural network theorists have used CQ as a core of extended models that have offered explanations of many of the grammar-respecting patterns of sequencing-errors observed in language production (e.g., Dell, Burger, & Svec, 1997; Hartley & Houghton, 1996).

The most sustained treatment of CQ in language generation is that in Ward (1994). Far from simply explaining how the ‘emergent choice’ that operates in CQ

models is compatible with grammar-respecting sequencing errors in language production, Ward argues that *only* emergent choice offers a basis for overcoming more traditional language generators' failures to mimic the 'flexible incremental generation' (FIG) exhibited in the real-time behavior of human speakers as they compose sentences 'on the fly'. Ward's FIG model combines CQ principles with principles inspired by *construction grammar* (e.g., Croft, 2001; Fillmore, 1988; Goldberg, 1995) to build a comprehensive connectionist model of grammatical sentence generation. The FIG algorithm is an iterated cycle: (1) each node of an input conceptualization is a source of activation to 'construction' nodes of various types, including words; (2) activation is allowed to flow freely through the structured network of nodes; (3) when the network settles (or is forced to make an output) the most highly activated word representation is selected and enacted; (4) any node or nodes of the input conceptualization that are expressed by the enacted word are inhibited, and activation levels are updated to represent the new current state; (5) steps 2–4 iterate until the conceptual content of the input has been expressed by the enacted word sequence. For the system to work well, the word plan that has the highest activation must be for a word which will be both syntactically and semantically correct if spoken as the next word in the utterance. This requirement is met, in part, by having the evolving activation level of a word be dynamically determined by the product of its semantic and syntactic inputs, not by their sum.

During the last two decades, CQ-compatible neural models have been explored in many further domains of learned serial behavior, including: eye movements (Grossberg & Kuperstein, 1986); phoneme sequences with repeating elements and phoneme coarticulation (Houghton, 1990); cursive handwriting (Bullock, Grossberg, & Mannes, 1993); working-memory storage of sequential inputs (Bradski, Carpenter, & Grossberg, 1994); word recognition and production (Grossberg, 1986; Gupta & MacWhinney, 1997); and melody learning and performance (Mannes, 1994; Page, 1994). These applications illustrate that the CQ model is highly extensible. For example, one of the advantages of CQ models' explicit parallel representation of sequential plans – an advantage unavailable to RNNs as such – is that these distributed representations can be learned and recalled via compressive and expansive coding operations. In the Sternberg task and the ISR tasks mentioned above, novel sequence information was provided to the performer. According to the CQ interpretation, performers hold a corresponding parallel representation for a few seconds in working memory (WM) before generating the sequence under the guidance of WM. However, Klapp (1996) and Verwey (1996), among others, showed that high numbers of practice trials with short fixed sequences leads to *disappearance* of the SLEL (Sternberg et al., 1978). This result can be explained by an augmented CQ model. Rhodes & Bullock (2002b) reported successful simulations of several sets of list learning and performance data, using a neural network in which the cerebellum, modeled as one substrate for procedural long-term memory (LTM), learns activation gradients over item nodes and rapidly recalls them into a normalized motor buffer (planning layer), which is a WM for action plans. The recall process is rapid because it entails *parallel loading of sequence chunks* into a WM from LTM. When the procedural LTM of a fixed sequence representation becomes strong enough

(due to extensive practice), it causes pre-selection of the first list-item within the CQ subsystem. Such pre-selection explains the practice-dependent disappearance of the SLEL. This hybrid cerebellar-CQ model's assumption that the cerebellum can load parallel sequence representations into a fronto-cortical buffer is supported by recent neuroanatomical tracing studies, which have discovered pathways that run from the dentate nuclei of the cerebellum, via the thalamus, to several fronto-cortical zones, including premotor cortex and the PFC (Dum & Strick, 2003). More generally, the hybrid model shows one way that the CQ model, which focuses on WM dynamics that support sequential performance, can interface with an LTM system that compressively learns and stores, and expansively recalls, oft-used sequences. Such a system may be critical for functions that require frequent re-use of subsequences, such as musical performance or language production. Note that loading such subsequences into a prefrontal (PFC) buffer makes their final expression in behavior subject to voluntary modulation in premotor and motor cortices, which are downstream of PFC. Such modulation of otherwise fixed sequences is critical both for emphatic aspects of speech and musical expressivity.

2.3. *Neurophysiological paradigms and results*

Sequence paradigms have been studied extensively in single cell neurophysiology experiments. A number of laboratories have carried out experiments in frontal lobe cortical areas including the supplementary motor area (SMA), the pre-supplementary motor area (pre-SMA), and the prefrontal cortex. In a series of experiments, Tanji and colleagues explored the neural representation, at the single cell level, of the elements of sequences. In their first study they used a task in which monkeys were trained to push buttons on a four-button touch pad in a particular order (Mushiake, Inase, & Tanji, 1990; Mushiake, Masahiko, & Tanji, 1991). This task was executed under two conditions, a visually guided condition and a memory condition. In the visually guided condition the sequence of movements was instructed by the sequential illumination of three buttons on the touch pad. After a GO signal the monkey executed the indicated sequence. In the memory condition, the animal first executed a series of six trials in which a fixed sequence was cued as in the visually guided trials. After the six visually guided trials were completed, the monkey had to execute the sequence following only a GO signal, without visual cueing. In the memory condition, neural activity related to sequence execution was prominent in the SMA, and a subset of neurons was preferentially active before the execution of a specific sequence. These neurons were not active before other memory-guided sequences, and were not active before the same sequence if executed in the visually guided condition.

In a second series of experiments (Shima, Mushiake, Saito, & Tanji, 1996; Shima & Tanji, 2000; Tanji & Shima, 1994) animals were trained on a serial order task in which they were required to carry out a sequence, the elements of which were one of three possible movements: a push, a pull or a turn of a manipulandum. The animals again carried out the task in visually guided and memory conditions and neurons were found which increased their activity before the execution of a preferred

sequence, only in the memory-guided condition, and not before other sequences. A second type of neuron was selective for the sequential position of the movement, independent of the specific movement executed. A third type of neuron was active selectively between two particular movements – for example, after a push and before a pull, but not after a push followed by a turn, or before a pull if it was preceded by a turn. Thus these neurons appeared to link two actions in a sequence.

Hikosaka and colleagues recorded in SMA and pre-SMA during monkeys' learning and performance of the $2 \times N$ task described above (Nakamura, Sakai, & Hikosaka, 1998). The responses of individual neurons were preferentially related to either the acquisition of new hypersets or the performance of previously learned hypersets. Seventy-eight out of 345 task related neurons responded more strongly to the performance of new hypersets than the performance of remembered hypersets. Of these 78 new preferring cells, 33 showed a learning-dependent decrease in activity; that is, their activity level was strongest at the beginning of the acquisition of a new hyperset, whereas 11 showed a learning-dependent increase in activity. Finally, 18 neurons responded preferentially to the production of a particular set within a learned hyperset. An analysis of the anatomical distribution of neurons showed that neurons related to learning of new sets were preferentially located in pre-SMA, whereas the SMA appeared to contain a roughly equal distribution of neurons related to new and learned sets – consistent with human imaging and primate reversible lesion results.

Clower & Alexander (1998) conducted a serial order experiment, in which identical movements were made as elements at multiple positions within several sequences. The task apparatus had four targets arrayed around a start hold circle. When one of the peripheral targets was cued, the monkey moved a cursor from the start hold circle to the peripheral target. After a hold period, one of the targets, positioned either clockwise or counterclockwise from the cued target, changed color briefly, instructing the direction of all subsequent movements. After the peripheral target changed color, the monkey moved in the direction which had been indicated previously. Another hold period followed and then the current target changed color, at which point the monkey moved to the next target maintaining the same direction of movement. The task continued in this way until the monkey had made three movements between the peripheral targets, visiting all four targets. By cueing different starting targets and different directions, each individual movement was produced at different serial positions. Thus the effect of the serial position of a movement could be assessed. Thirty-nine percent of SMA neurons recorded, and 71% of pre-SMA neurons showed an effect of serial position on their response.

In general, these experiments in the SMA and pre-SMA found that a neuron's firing rate will change when a particular movement is being executed at different positions within a sequence, or within different sequences. While these experiments have provided valuable data on the cortical representation of the elements of sequences, they have provided little definitive support for, or insight into, possible mechanisms underlying the production of the elements of a sequence in their correct order. As discussed above, associative chaining and parallel response activation have been proposed as mechanisms by which the elements of a sequence can be produced in their

correct order. These theories make different predictions about their associated neuro-physiological signals. Specifically, associative chaining predicts that only a single element of a sequence will be active at any point in time, and the activation of this element will ‘cause’ the next element to become active. Contrary to the predictions of the associative chaining model, parallel response activation models, such as CQ, predict that the elements of the sequence will be represented simultaneously, and *in parallel*. Thus, even before the sequence is executed all the elements of the sequence will be simultaneously activated.

Averbeck et al. (2002), Averbeck, Chafee, et al. (2003), and Averbeck, Crowe, et al. (2003) have reported results from an experiment in which monkeys were trained to use a prescribed stroke sequence to draw a set of geometric shapes (prescribed sequence figure drawing, PSFD), including a triangle, square, trapezoid, and upside down triangle. In their experiment, the monkeys began a trial by maintaining a joystick controlled cursor in a start hold circle for 1 s. After this 1 s hold period, a static template (geometric form) appeared on the right half of the screen, and the monkey was free to draw on the left half. If the monkey executed a complete drawing trajectory, while keeping the cursor within (non-visible) ‘corridors’ that defined acceptable form, a juice reward was delivered. Shapes were drawn in blocks of consecutive trials of the same shape. This regularity, and the 1 s hold period, allowed the monkey to anticipate, and prepare to draw, the appropriate shape in the subsequent trial, on

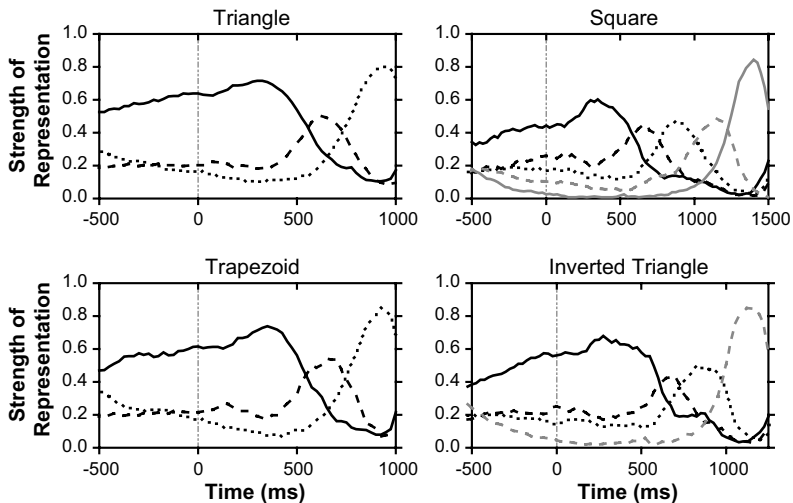


Fig. 5. Plots of strength of segment representation for four different shapes vs. time (from Averbeck et al., 2002, Fig. 2). Time 0 (vertical broken grey line) indicates the onset of the template which permitted drawing to commence. Consistent with competitive queuing (CQ) models, the plots show parallel representation of segments before initiation of copying; rank order of strength of representation before copying corresponds to the serial position of the segment in the series; and the rank order evolves during the drawing to maintain the serial position code. Line color/pattern corresponds to segments as follows: solid black, segment 1; dashed black, segment 2; dotted black, segment 3; dashed grey, segment 4; solid grey, segment 5 (but not all segments are defined for all shapes).

all trials except the first trial of a block. Analysis of the acceleration profiles of the monkey's hand movements showed that the continuous trajectory was composed of a sequence of individual segments (Averbeck, Chafee, et al., 2003). While the monkeys carried out this task, ensembles of individually isolated single neurons were recorded in the prefrontal cortex. Neural activity patterns were defined, based upon the average ensemble neural responses which occurred during the drawing of individual segments of the geometric shapes. These activity patterns were shown to be highly predictive of the shape segment being drawn (Averbeck, Crowe, et al., 2003), and therefore could be considered neural correlates of each segment of the shape. When the neural activity during the hold period was analyzed, a parallel representation of these activity patterns was found (see Fig. 5). Furthermore, the relative strength of the representation of each segment also predicted the serial position of the segment, such that prior to the execution of the sequence, the first segment had the strongest representation, the second had the second strongest representation, etc. – a 'primacy gradient' as defined earlier. This parallel representation continued to unfold during the execution of the sequence of shape segments. After a segment was executed, its representation decreased, and the subsequent segment became the most strongly activated. Thus, this study provided the first neural evidence for a mechanism which could order the elements of a sequence, strongly supporting the CQ class of models reviewed above.

3. A contemporary (partial) synthesis: The N-STREAMS model

As the foregoing attests, learning and production of serial movements have received much attention from psychological and neuroscience experimentalists as well as modelers. Numerous models have attempted to address relatively specific parts of the data presented above or have explored a single learning mechanism as a basis for serial learning. Very few models have addressed data and neuroanatomical constraints simultaneously. Given the complexity of the picture painted by the data above, it is unlikely that a single, homogeneous mechanism will ultimately account for serial phenomena in general. A new neural network theory (Rhodes, 2000; Rhodes & Bullock, 2002b), known as N-STREAMS (an acronym for Neural Substrates That Rehearse, Encode, And Memorize Sequences), specifies interactions among several distinct bases for serial movement learning and performance. In addressing timing data as reported by Sternberg et al. (1978), Klapp (1995), and Verwey (1996), the model accounts for temporal characteristics of serial behavior performance both early and late in learning, and provides a unified treatment of changes that occur along the learning continuum. Key data properties that the model exhibits are: (1) a SLEL early in practice that disappears with extended practice; (2) a ratio effect, i.e., the pattern of long latency followed by markedly shorter inter-response intervals (IRIs) for non-initial sequence elements, under two conditions: either with foreknowledge of a novel sequence to be produced and adequate opportunity to prepare for its execution, or without preparation but after significant amounts of practice; (3) a SLER that does not disappear with practice; and (4) a serial position

dependence of IRIs that disappears after extended practice. The model also exhibits word-length effects such as those from ISR tasks (e.g., Cowan, Wood, & Borne, 1994; Hulme, Newton, Cowan, Stuart, & Brown, 1999) and the patterns of errors are consistent with those in the data. As already noted, the primacy model of Page & Norris (1998) uses a noisy CQ mechanism with a primacy gradient (in common with the N-STREAMS model) to very effectively model the patterns of error in sequence recall from short-term memory. The two models can be considered close relatives.

The major elements of the N-STREAMS model (see Fig. 6) are: (1) a fronto-cortical gradient-based representation of serial order that provides a sequence production buffer and CQ; (2) a working-memory mechanism capable of producing a suitable gradient from an input stream of items (representing a sequence) which also allows for comparison of sequences recalled from long-term memory with a representation currently being constructed on the basis of external stimulus presentation; (3) compressive chunk encoding for cortical sequence learning that enables efficient memorization of short sequences of items and facilitates recall for production or recognition purposes; and (4) a cerebellum-based learning module that learns both sequence chunks and individual inter-response transitions. As noted above for CQ in general, the gradient representation uses the relative activation level of primed items as an implicit code of serial order. Through practice, the cerebellar learning mechanism learns to anticipate and preempt slower cortical loading of the appropriate gradient into the frontal production buffer as well as to speed up the execution of individual responses within the sequence. These major components, along with others included within the model, are compatible with neuroanatomical constraints and with the major trends emerging from neurophysiological, clinical, and brain imaging investigations of learning and performance of serial movements. The model in particular highlights the functional significance of projections from the deep cerebellar nuclei to the frontal cortex via the motor thalamus (e.g., Dum & Strick, 2003).

Production of novel, preloaded sequences is accomplished by the ‘execution module’ of N-STREAMS. Although sharing gradient-based representation and competitive selection of individual items with the class of earlier CQ models of serial movement production (e.g., Grossberg, 1978a, 1978b; Houghton, 1990; Page & Norris, 1998), this module resolves many implementation problems not immediately evident within these earlier models given their algorithmic or difference equation specification. Only when implemented within a self-contained differential equation framework do many of these problems become obvious – and thus require resolution. Examples of such problems include effective deletion of selected items from the gradient buffer by feedback signals and prevention of premature selection of a subsequent item (before execution of the item currently being performed). This module also incorporates an automatic, but competitive, gain control system that governs overall function during the performance of a sequence. In the model, gain switches between the gradient buffer and the components of the module responsible for actual execution of the selected item. This competition for gain assists in solving the problem of effectively deleting items from the buffer once they have been selected for performance. It also embodies the type of working-memory dynamics that

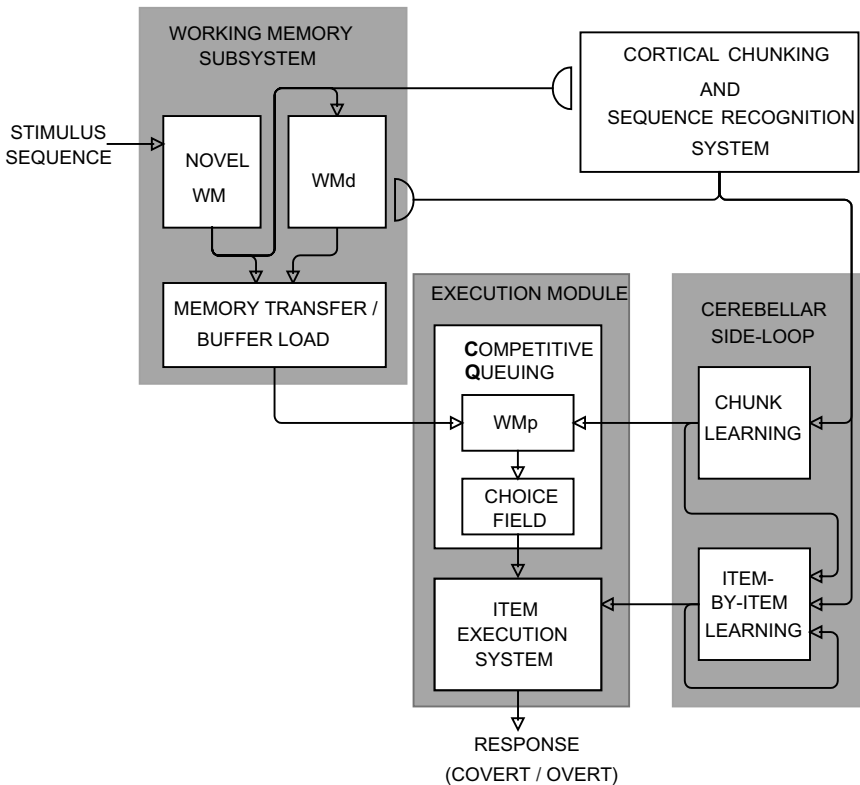


Fig. 6. Macro-circuit illustrating the global architecture of the N-STREAMS model. Only major components and links between them are depicted here, but the associated simulation model shows how these major parts can be fully implemented as neural networks. The name N-STREAMS is an acronym for Neural Substrates That Rehearse, Encode, and Memorize Sequences. It captures the hypothesis that several substrates capable of sequence representation combine their outputs to compose the stream of behavioral outputs. At the core of N-STREAMS is a competitive queuing (CQ) system (see Fig. 4). This includes a plan working memory (Wmp) and a choice field capable of choosing (for performance) the most active remaining plan in Wmp. The CQ core is augmented by a declarative WM (WMd) and a closely-associated adaptive cortical chunking system, which learns to recognize and recall the parallel sequence representations that pass through the WMd. Sites of learning between WMd and the cortical chunking subsystem are shown as semi-circles. Additional sites of learning are located in the other major augmentation of the CQ core, namely the cerebellar side loop. The text emphasizes the upper cerebellar side loop, which generates a parallel output to the CQ plan layer (Wmp). The lower cerebellar side loop illustrates the cerebellum's output-to-input recurrence, which allows the N-STREAMS model to utilize item-to-item associative links as a subsidiary sequencing mechanism. Adapted with permission from Rhodes and Bullock (2002b).

Cowan (1994) and Page & Norris (1998) proposed to explain the word-length effect. In the model, the working-memory representation (the primacy gradient) decays during actual performance of an item and is then refreshed by searching that representation for the subsequent item (see Fig. 7, top). As noted above, this is an area of continuing debate, and there are alternative explanations for the word-length effect,

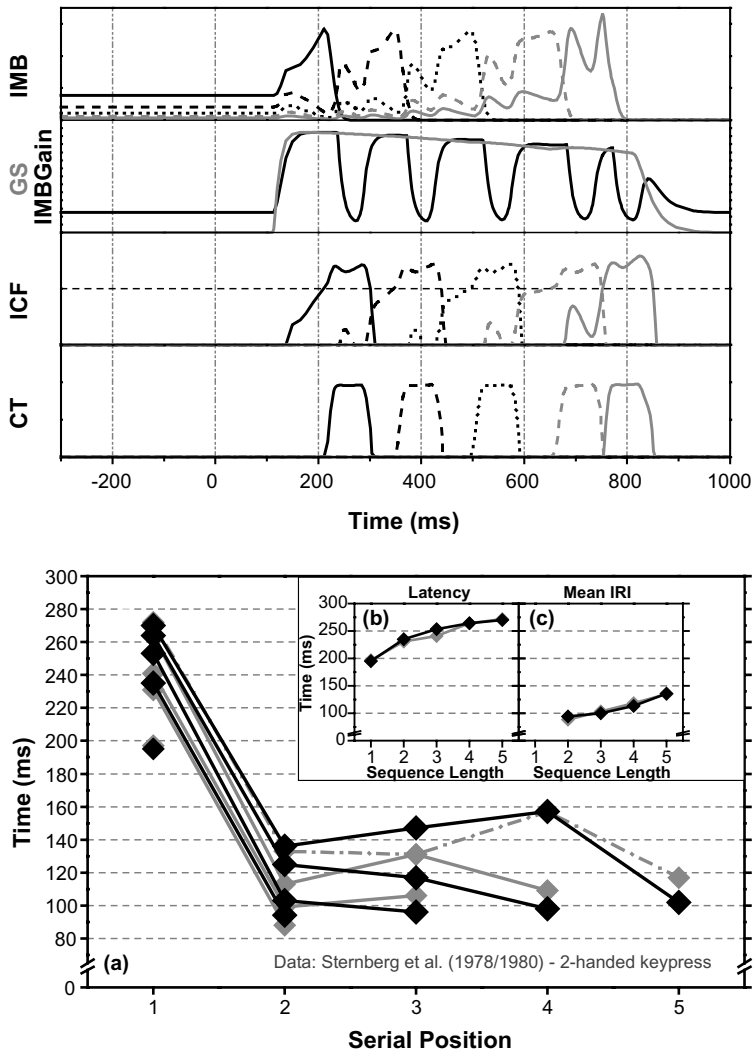


Fig. 7. Top: Timecourse of modeled cell activities comprising the competitive queuing component within the execution module of N-STREAMS during performance of a 5-item sequence (using line color/pattern convention of Fig. 5). The traces in the IMB (Item Motor Buffer) panel should be compared with those of the ‘Square’ panel of Fig. 5. Bottom: Execution module performance (colored black) compared with the Sternberg et al. (1978) 2-handed typing task data (colored grey).

some based on a different conception of decay (e.g., Page & Norris, 1998), and some that eschew decay entirely (e.g., Neath & Nairne, 1995). Irrespective of how this issue resolves, by itself, the N-STREAMS model’s execution module does exhibit the human operating characteristics – the SLEL and SLER – evident in RT studies of performance of novel sequences (e.g., Sternberg et al., 1978; see Fig. 7, bottom).

Following the lead of [Bradski et al. \(1994\)](#), a working-memory submodule is incorporated to enable the N-STREAMS to construct a gradient representation as the model is presented with a series of input items – as would be the case in any task requiring performance of a novel sequence, such as a phone number. But in N-STREAMS, this gradient can also be voluntarily transferred to the buffer of the execution module. This transfer process, a key constituent of voluntary preparation, facilitates learning of the sequence in two other parts of the model. The first is a cortical chunking component capable of learning a compressive representation of the loaded sequence. This form of learning could (at least in part) constitute a form of explicit learning. It allows subsequent recognition, and recall of the sequence into the working-memory subsystem, when initial items of the same sequence are presented at a later time. This recall constitutes a ‘best guess’ (from currently known sequences) as to which sequence an incoming stream of stimuli may represent. Further accumulation of evidence, on the basis of presentation of additional stimuli, serves to either confirm or disconfirm this hypothesis.

The transfer of the sequence from working memory to the production buffer within the execution module also provides a teaching signal to the cerebellar module. This teaching signal causes the latter to learn its own gradient representation of the sequence. After sufficient practice, presentation of the initial item(s) of a learned sequence causes the recognition component to provide the cerebellar module with a specific contextual input. Appearance of this input triggers the cerebellar module to rapidly instate its learned gradient representation of the sequence into the fronto-cortical production buffer. This trans-cerebellar loading of the frontal buffer occurs much more quickly than loading that utilizes the working memory. Such speeded loading of well-learned sequences can explain learning-dependent changes in the latency to produce the first item of a sequence, notably the loss of the SLEL for specific, highly practiced, sequences.

Intra-sequence transitions, from earlier to later items, are also learned by the cerebellar module. This item-by-item learning marks a second role within the overall N-STREAMS model for the cerebellar side loop. The differentiation between roles is solely based upon the input and output connections to and from the cerebellar circuits – as is the case in vivo. With extensive practice of a sequence, cerebellar learning reduces the latency between items and speeds up production of the entire sequence. Doing so relies upon the adaptive timing competence that is known to be provided by the cerebellar cortex (e.g., [Fiala, Grossberg, & Bullock, 1996](#); [Perrett, Ruiz, & Mauk, 1993](#)). The necessity of embedding this competence within the context of sequence performance emphasized the issue of scalability and the importance of the recurrent nature of the cerebellar circuitry, and resulted in the development of a new model of cerebellar adaptive timing and sequencing. This Recurrent Slide and Latch (RSL) model is introduced and documented in [Rhodes & Bullock \(2002a\)](#). There, it is noted that timing and sequencing operations requiring entire cell populations in alternative models require only a few cells in the RSL module. The learning that occurs within this cerebellar module is a form of procedural learning.

A key feature of the dynamical N-STREAMS model is stable, self-regulated interaction, from initial to late stages of practice, among the various components of the

model. The theory explains how each component contributes different competencies, all of which appear necessary when attempting to explain the brain system responsible for serial learning and production. The involvement of multiple substrates is now well established (e.g., Hikosaka et al., 1999; Sakai et al., 1998 as noted above). Of special interest is the model's incorporation of parallel and serial representations within a consistent and unified framework, because it allows a reconciliation of mechanisms previously treated as exclusive alternatives. The gradient buffer, chunking apparatus, and the first of the cerebellar module roles are fundamentally parallel (where sequence items are temporally co-active); the second cerebellar role, with its exploitation of recurrence, is fundamentally a serial mechanism (whereby only a single item is active at any given time). The latter is typical of sequence learning and production models in which the sequence is not explicitly represented, but instead is recovered only when the system runs – as in the RNNs referred to above (e.g., Cleeremans & McLelland, 1991; Dominey & Arbib, 1992; Elman, 1990; Jordan, 1989). Thus, the current N-STREAMS model hypothesizes a distinct role for each of two major classes of mechanism previously proposed to explain serial organization in learning and performance.

Although the model has not yet been applied beyond the domain of button pressing tasks, it is extensible to cover many types of performance in which subjects learn stable sequences defined over finite sets of items. Included here are linguistic performances, such as typing, handwriting and speech production. The structure of the model makes it compatible with prior cognitive proposals that emphasize how chunking maximizes effective use of working memory and output buffers that have a severely limited (e.g., 4–7 item) capacity. For handwriting, the most compatible treatments are dynamic neural network models that generate cursive forms via overlapped readout of a small number of discrete linear strokes represented in a motor buffer (e.g., Bullock et al., 1993; Contreras-Vidal, Poluha, Teulings, & Stelmach, 1998).

The compressive cortical chunking competence of N-STREAMS provides a basis from which to begin to address issues related to sequencing chunks, as have recently been elucidated by Verwey (2001), for example. The interaction between preparation of a forthcoming sequence and execution of the present sequence represents fertile ground for the continued development of the N-STREAMS theoretical framework. Another interesting avenue for development would be the early learning phase of the $2 \times N$ task. Here, the working-memory component of the current N-STREAMS formulation would provide a substrate for additional development to enable a competence for trial and error with search. The search aspect would be facilitated by a suitable memorization and recognition mechanism that could be accomplished by the current chunk learning component of the model. N-STREAMS features intra-sequence transition (item-by-item) learning within a cerebellar side loop. This aspect of the model is a ready substrate for the type of learning indicated by the SRT tasks described earlier. The specific nature of the cerebellar learning taking place in the N-STREAMS model suggests that this mechanism would result in sequence learning under conditions where such learning occurs in experiments (e.g., with stable inter-stimulus intervals) whereas conditions preventing learning in experiments (such as

random inter-stimulus intervals) would also prevent learning in the N-STREAMS model. An open question is the range of tasks that the N-STREAMS model can explain. In tasks where the subject is aware of the sequence and chunking is known to occur, a cerebellar contribution to LTM for learned sequences has been demonstrated (e.g., Lu et al., 1998). However, one recent neuroimaging report has challenged the idea that the cerebellum is also involved in the implicit learning that occurs in the SRT (serial reaction time) task, in which subjects often show evidence of sequence learning despite having no awareness that a non-random sequence has been presented and learned. In particular, Seidler et al. (2002) showed that throughout an initial learning phase that included a distractor task performed in parallel with the SRT task, there was no evidence of learning in performance, and also no evidence of cerebellar activations. Upon removal of the distractor task, performance suddenly improved, and cerebellar activations suddenly became detectable. They interpreted this correlation to mean that removal of the distractor task suddenly enabled previously masked cortical learning to gain expression *via the cerebellum* and thereby immediately enhance performance. However, correlation does not imply causation. The N-STREAMS theory is fully consistent with the Seidler et al. (2002) data, but it offers a strikingly different interpretation: removal of the distractor task allows cortical chunk learning to finally gain access to the WMp, via which such learning immediately begins to have an impact on performance (even *without* a cerebellar assist to performance). Simultaneously, the transfers from WMd to WMp generate teaching signals for cerebellar learning, and it is these teaching signals that cause the cerebellar activations observed in the neuroimaging study. This prediction remains to be tested, but it can be seen that the N-STREAMS framework, even in its early stage of development, has the potential to span many of the paradigms and results presented earlier in this paper.

4. Conclusions

The present discussion of research paradigms, tasks and models of skilled sequential motor behavior indicates that people have the capacity to control short sequences as chunks whose elements can be treated collectively, e.g., activated in parallel, during cognitive operations. Such collective treatment may be a necessary condition for hierarchical control, which is further suggested by many of the data and models reviewed. In such hierarchical control, short segments can be processed automatically, in the sense that their initiation and execution need not require shifts of attention and deliberative executive control, and need not burden the kind of short-term memory required for recoding between alternative representations. Most likely, these segments are coded in a task-specific way that facilitates rapid processing.

On the other hand, there is ample reason to believe that a kind of working memory can mediate performance of even well-learned short sequences. The need for continuing working-memory involvement makes sense from several perspectives. First, we know that humans are able to modulate the performance of very well-

learned short sequences at will, as when a teacher greatly elongates a spoken syllable to meet some transient communicative goal. Second, when a chunk is unpacked into its constituent representations, the activation of the later elements must be sustained until the earlier elements have been performed. The length of time needed can be so short as to hardly require working memory, but for one reason or another, it can also be long enough to require the sustained activation provided by the working-memory system long associated with the pre-frontal cortex.

Within hierarchical control models, a key issue is whether higher and lower-level control processes may be carried out simultaneously. In general, simultaneous processing is to be preferred because it can greatly speed system operation. Above the chunk level, control involves information and decisions regarding which next chunk to initiate, and recent research suggests that such control operates simultaneously with lower-level operations. From a processing point of view, this demonstrates that independent processors are responsible for low and high level control, most likely tapping different knowledge bases. On the other hand, there are also indications from the ISR literature that during short sequence performance, there may be at least a brief alternation between two iterated phases: launching item execution and running the competition to choose the next item for execution.

A promising recent development is the emergence of adaptive neural network models that respect neuroanatomical and neurophysiological constraints and that are applicable to sequence learning in addition to other tasks. One problem with these models is that they tend to be enormously complex, with much more internal structure and far more parameters than traditional mathematical models in the psychology of sequence learning. For biological realism, these models must be complex. It is therefore incumbent on the modelers to demonstrate that their models are competent to explain a much wider range of data than has been targeted by most traditional models in psychology. For example, a model should be able to explain real-time performance measures (e.g., latencies) and error patterns as they evolve across early, middle and asymptotic phases of task learning and performance.

Acknowledgments

BJR was supported in part by DARPA/ONR N00014-95-1-0409. DB was supported in part by NIH R01 DC02852. BBA was supported by United States Public Health Service grant NS17413. MPAP gratefully acknowledges the support of the UK Biotechnology and Biosciences Research Council via Grant 310/S15906.

References

- Adams, J. A. (1984). Learning of movement sequences. *Psychological Bulletin*, *96*, 3–28.
- Adi-Japha, E., & Freeman, N. H. (2000). Regulation of division of labour between cognitive systems controlling action. *Cognition*, *76*, 1–11.
- Anson, J. G. (1982). Memory drum theory: Alternative tests and explanations for the complexity effects on simple reaction time. *Journal of Motor Behavior*, *14*, 228–246.

- Averbeck, B. B., Chafee, M. V., Crowe, D. A., & Georgopoulos, A. P. (2002). Parallel processing of serial movements in prefrontal cortex. *Proceedings of the National Academy of Sciences*, *99*, 13172–13177.
- Averbeck, B. B., Chafee, M. V., Crowe, D. A., & Georgopoulos, A. P. (2003). Neural activity in prefrontal cortex during copying geometrical shapes. I. Single cell studies. *Experimental Brain Research*, *150*, 127–141.
- Averbeck, B. B., Crowe, D. A., Chafee, M. V., & Georgopoulos, A. P. (2003). Neural activity in macaque prefrontal cortex during copying geometrical shapes. II. Decoding shape segments from neural ensembles. *Experimental Brain Research*, *150*, 142–153.
- Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory? *Quarterly Journal of Experimental Psychology*, *20*, 249–263.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behaviour*, *14*, 575–589.
- Baird, H. P., Noble, M., & Fitts, P. M. (1954). Extra-task performance as a measure of learning a primary task. *Journal of Experimental Psychology*, *48*, 298–302.
- Bain, A. (1868). *The sense and the intellect* (3rd ed.). London: Longmans, Green.
- Basso, M. A., & Wurtz, R. H. (1998). Modulation of neuronal activity in superior colliculus by changes in target probability. *Journal of Neuroscience*, *18*, 7519–7534.
- Boardman, I. S. (1995). *Neural network models of temporal processing in speech perception and motor control*. Unpublished doctoral dissertation, Department of Cognitive and Neural Systems, Boston University, Boston.
- Boardman, I. S., & Bullock, D. (1991). A neural network model of serial order recall from short-term memory. *Proceedings of the International Joint Conference on Neural Networks (Seattle)* (Vol. II, pp. 879–884). Piscataway, NJ: IEEE Service Center.
- Book, W. F. (1908). *The psychology of skill*. Missoula: Montana Press.
- Bradski, G., Carpenter, G. A., & Grossberg, S. (1994). STORE working memory networks for storage and recall of arbitrary temporal sequences. *Biological Cybernetics*, *71*, 469–480.
- Broadbent, D. E. (1977). Levels, hierarchies, and the locus of control. *Quarterly Journal of Experimental Psychology*, *29*, 181–201.
- Brown, T. L., & Carr, T. H. (1989). Automaticity in skill acquisition: Mechanisms for reducing interference in concurrent performance. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 686–700.
- Bullock, D., Grossberg, S., & Mannes, C. (1993). A neural network model for cursive script production. *Biological Cybernetics*, *70*, 15–28.
- Bullock, D., Liederman, J., & Todorovic, D. (1987). Reconciling stable asymmetry with recovery of function: An adaptive systems perspective on functional plasticity. *Child Development*, *58*, 689–697.
- Bullock, D., & Rhodes, B. J. (2003). Competitive queuing for planning and serial performance. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (2nd ed., pp. 241–244). Cambridge, MA: MIT Press.
- Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language*, *31*, 429–460.
- Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word-length effects in span tasks. *Quarterly Journal of Experimental Psychology*, *45A*, 177–192.
- Christina, R. W., Fischman, M. G., Vercruyssen, M. J. P., & Anson, J. G. (1982). Simple reaction time as a function of response complexity: Memory drum revisited. *Journal of Motor Behavior*, *14*, 301–321.
- Cisek, P., & Kalaska, J. F. (2002). Simultaneous encoding of multiple potential reach directions in dorsal premotor cortex. *Journal of Neurophysiology*, *87*, 1149–1154.
- Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, *2*, 406–416.
- Cleeremans, A., & McLelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, *120*, 235–253.

- Clower, W. T., & Alexander, G. E. (1998). Movement sequence-related activity reflecting numerical order of components in supplementary and presupplementary motor areas. *Journal of Neurophysiology*, *80*, 1562–1566.
- Cohen, A., Ivry, R. B., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 17–30.
- Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behaviour*, *15*, 17–31.
- Conrad, R. (1958). Accuracy of recall using keyset and telephone dial, and effect of a prefix digit. *Journal of Applied Psychology*, *42*, 285–288.
- Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, *55*, 75–84.
- Contreras-Vidal, J. L., Poluha, P., Teulings, H. L., & Stelmach, G. E. (1998). Neural dynamics of short and medium-term motor control effects of levodopa therapy in Parkinson's disease. *Artificial Intelligence in Medicine*, *13*, 57–79.
- Conway, C. C., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, *5*, 539–546.
- Cowan, N. (1994). Mechanisms of verbal short-term memory. *Current Directions in Psychological Science*, *3*, 185–189.
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.
- Cowan, N., Wood, N. L., & Borne, D. N. (1994). Reconfirmation of the short-term storage concept. *Psychological Science*, *5*, 103–106.
- Croft, W. (2001). *Radical construction grammar*. New York: Oxford.
- Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*, 123–147.
- Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin and Review*, *8*, 343–350.
- Destrebecqz, A., Peigneux, P., Laureys, S., Degueldre, C., Del Fiore, G., Aerts, J., et al. (2003). Cerebral correlates of explicit sequence learning. *Cognitive Brain Research*, *16*, 391–398.
- Dominey, P. F., & Arbib, M. A. (1992). A cortico-subcortical model for generation of spatially accurate sequential saccades. *Cerebral Cortex*, *2*, 153–175.
- Doyon, J., Gaudreau, D., Laforce, Jr., R., Castonguay, M., Bedard, P. J., Bedard, F., et al. (1997). Role of the striatum, cerebellum, and frontal lobes in the learning of a visuomotor sequence. *Brain and Cognition*, *34*, 218–245.
- Dum, R. P., & Strick, P. L. (2003). An unfolded map of the cerebellar dentate nucleus and its projections to the cerebral cortex. *Journal of Neurophysiology*, *89*, 634–639.
- Eimer, M., Goschke, T., Schlaghecken, F., & Stormer, B. (1996). Explicit and implicit learning of event sequences: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 970–987.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.
- Ferraro, F. R., Balota, D. A., & Connor, L. T. (1993). Implicit memory and the formation of new associations in nondemented Parkinson's disease individuals and individuals with senile dementia of the Alzheimer's type: A serial reaction time (SRT) investigation. *Brain and Cognition*, *21*, 163–180.
- Fiala, J. C., Grossberg, S., & Bullock, D. (1996). Metabotropic glutamate receptor activation in cerebellar Purkinje cells as substrate for adaptive timing of the classically conditioned eye-blink response. *Journal of Neuroscience*, *16*, 3760–3774.
- Fillmore, C. J. (1988). The mechanisms of construction grammar. In *Berkeley Linguistics Society, Proceedings of the Fourteenth Annual Meeting* (pp. 35–55).
- Freeman, F. N. (1907). Preliminary experiments on writing reactions. *Psychological Monographs*, *8*, 301–333.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*, 20–25.

- Gordon, P. C., & Meyer, D. E. (1987). Control of serial order in rapidly spoken syllable sequences. *Journal of Memory and Language*, 26, 300–321.
- Grafton, S. T., Hazeltine, E., & Ivry, R. B. (1995). Functional mapping of sequence learning in normal humans. *Journal of Cognitive Neuroscience*, 7, 497–510.
- Grossberg, S. (1978a). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology* (Vol. 5, pp. 233–374). New York: Academic Press.
- Grossberg, S. (1978b). Behavioral contrast in short term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology*, 17, 199–219.
- Grossberg, S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines, Volume 1: Speech perception* (pp. 187–294). New York: Academic Press.
- Grossberg, S., & Kuperstein, M. (1986). *Neural dynamics of adaptive sensory-motor control*. Amsterdam: Elsevier/North Holland.
- Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language*, 59, 267–333.
- Hartley, T. A., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language*, 35, 1–31.
- Hazeltine, E., Grafton, S. T., & Ivry, R. (1997). Attention and stimulus characteristics determine the locus of motor sequence encoding: A PET study. *Brain*, 120, 123–140.
- Helmholtz, H. von (1866). *Handbuch der physiologischen Optik*. Leipzig: Voss.
- Henry, F. M. (1980). Use of simple reaction time in motor programming studies: A reply to Klapp, Wyatt, and Lingo. *Journal of Motor Behavior*, 12, 163–168.
- Henry, F. M., & Rogers, D. E. (1960). Increased response latency for complicated movements and a 'memory drum' theory of neuromotor reaction. *Research Quarterly*, 31, 448–458.
- Henson, R. N. A. (1996). *Short-term memory for serial order*. Unpublished doctoral dissertation, University of Cambridge.
- 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. *Quarterly Journal of Experimental Psychology*, 49A, 80–115.
- Heuer, H. (1988). Testing the invariance of relative timing: Comment on Gentner (1987). *Psychological Review*, 95, 552–557.
- Heuer, H., Spijkers, W., Kieswetter, E., & Schmidtke, V. (1998). Effects of sleep loss, time of day, and extended mental work on implicit and explicit learning of sequences. *Journal of Experimental Psychology: Applied*, 4, 139–162.
- Hikosaka, O., Nakahara, H., Rand, M. K., Sakai, K., Lu, X., Nakamura, K., et al. (1999). Parallel neural networks for learning sequential procedures. *Trends in Neurosciences*, 22, 464–471.
- Hikosaka, O., Rand, M. K., Miyachi, S., & Miyashita, K. (1995). Learning of sequential movements in the monkey: Process of learning and retention of memory. *Journal of Neurophysiology*, 74, 1652–1661.
- Hikosaka, O., Sakai, K., Miyauchi, S., Takino, R., Sasaki, Y., & Putz, B. (1996). Activation of human presupplementary motor area in learning of sequential procedures: A functional MRI study. *Journal of Neurophysiology*, 76, 617–621.
- Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current research in natural language generation* (pp. 287–319). London: Academic Press.
- Hulme, C., Newton, P., Cowan, N., Stuart, G., & Brown, G. (1999). Think before you speak: Pauses, memory search, and trace reintegration processes in verbal memory span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 447–463.
- Hulstijn, W., & van Galen, G. P. (1988). Levels of motor programming in writing familiar and unfamiliar symbols. In A. M. Colley & J. R. Beech (Eds.), *Cognition and action in skilled behaviour* (pp. 65–85). Amsterdam: North-Holland.
- Inhoff, A. W., Diener, H. C., Rafal, R. D., & Ivry, R. (1989). The role of cerebellar structures in the execution of serial movements. *Brain*, 112, 565–581.

- Inhoff, A. W., & Rafal, R. D. (1990). Cerebellar structures and the programming of movement sequences. *Behavioural Neurology*, 3, 87–97.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513–541.
- James, W. (1890). *The principles of psychology* (Vol. 1). New York: Holt.
- Jimenez, L., Mendez, C., & Cleeremans, A. (1996a). Comparing direct and indirect measures of sequences learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 948–969.
- Jimenez, L., Mendez, C., & Cleeremans, A. (1996b). Measures of awareness and of sequence knowledge. In *Psyche: An interdisciplinary journal of research on consciousness* (On-line). Retrieved from <http://psyche.cs.monash.edu.au>.
- Jones, D., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 369–381.
- Jones, D., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *Quarterly Journal of Experimental Psychology*, 44A, 645–669.
- Jordan, M. I. (1989). Serial order: A parallel distributed processing approach. In J. L. Elman & D. E. Rumelhart (Eds.), *Advances in connectionist theory: Speech* (pp. 214–249). Hillsdale, NJ: Erlbaum.
- Keele, S. W. (1968). Movement control in skilled motor performance. *Psychological Bulletin*, 70, 387–403.
- Keele, S. W., Ivry, R. B., Hazeltine, E., Mayr, U., & Heuer, H. (2003). The cognitive and neural architecture of sequence representations. *Psychological Review*, 110, 316–339.
- Klapp, S. T. (1995). Motor response programming during simple and choice reaction time: The role of practice. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 1015–1027.
- Klapp, S. T. (1996). Reaction time analysis of central motor control. In H. Zelaznik (Ed.), *Advances in motor learning and control*. Champaign, IL: Human Kinetics Press.
- Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, 302, 1181–1185.
- Kornbrot, D. E. (1989). Organization of keying skills: The effect of motor complexity and number of units. *Acta Psychologica*, 70, 19–41.
- Lashley, K. S. (1917). The accuracy of movement in the absence of excitation from the moving organ. *American Journal of Physiology*, 43, 169–194.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley (Reprinted in *The neuropsychology of Lashley*, pp. 506–528, by Beach, F.A., Hebb, D.O., Morgan, C.T., & Nissen, H.W. (Eds.), 1960, New York: McGraw-Hill).
- Lovatt, P., Avons, S. E., & Masterson, J. (2000). The word-length effect and disyllabic words. *The Quarterly Journal of Experimental Psychology*, 53A, 1–22.
- Lovatt, P., Avons, S. E., & Masterson, J. (2002). Output decay in immediate serial recall: Speech time revisited. *Journal of Memory and Language*, 46, 227–243.
- Lu, X., Hikosaka, O., & Miyachi, S. (1998). Role of monkey cerebellar nuclei in skill for sequential movement. *Journal of Neurophysiology*, 79, 2245–2254.
- MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review*, 89, 483–506.
- MacKenzie, C. L., & van Eerd, D. L. (1990). Rhythmic precision in the performance of piano scales: Motor psychophysics and motor programming. In M. Jeannerod (Ed.), *Attention and performance XIII* (pp. 375–408). Hillsdale, NJ: Erlbaum.
- MacWhinney, B. (2003). Language acquisition. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (2nd ed., pp. 600–603). Cambridge, MA: MIT Press.
- Mannes, C. (1994). *Neural network models of serial order and handwriting movement generation*. Unpublished doctoral dissertation, Department of Cognitive and Neural Systems, Boston University.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.
- Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt.

- Miyachi, S., Hikosaka, O., Miyashita, K., Karadi, Z., & Rand, M. K. (1997). Differential roles of monkey striatum in learning of sequential hand movement. *Experimental Brain Research*, *115*, 1–5.
- Miyashita, K., Rand, M. K., Miyachi, S., & Hikosaka, O. (1996). Anticipatory saccades in sequential procedural learning in monkeys. *Journal of Neurophysiology*, *76*, 1361–1366.
- Mushiake, H., Inase, M., & Tanji, J. (1990). Selective coding of motor sequence in the supplementary motor area of the monkey cerebral cortex. *Experimental Brain Research*, *82*, 208–210.
- Mushiake, H., Masahiko, I., & Tanji, J. (1991). Neuronal activity in the primate premotor, supplementary, and precentral motor cortex during visually guided and internally determined sequential movements. *Journal of Neurophysiology*, *66*, 705–718.
- Nakamura, K., Sakai, K., & Hikosaka, O. (1998). Neuronal activity in medial frontal cortex during learning of sequential procedures. *Journal of Neurophysiology*, *80*, 2671–2687.
- Nakamura, K., Sakai, K., & Hikosaka, O. (1999). Effects of local inactivation of monkey medial frontal cortex in learning of sequential procedures. *Journal of Neurophysiology*, *82*, 1063–1068.
- Neath, I., & Nairne, J. S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin and Review*, *2*, 429–441.
- Newell, K. M. (1978). Some issues on action plans. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 41–54). New York: Academic Press.
- Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1–32.
- Norris, D., Page, M. P. A., & Baddeley, A. D. (1994, July). *Serial recall: It's all in the representations*. Paper presented at the International Conference on Working Memory, Cambridge, UK.
- Page, M. P. A. (1994). Modeling the perception of musical sequences with self-organizing neural networks. *Connection Science*, *6*, 223–246.
- Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*, 761–781.
- Pellizzer, G., & Hedges, J. H. (2003). Motor planning: Effect of directional uncertainty with discrete spatial cues. *Experimental Brain Research*, *150*, 276–289.
- Perrett, S. P., Ruiz, B. P., & Mauk, M. D. (1993). Cerebellar cortex lesions disrupt learning-dependent timing of conditioned eyelid responses. *Journal of Neuroscience*, *13*, 1708–1718.
- Perruchet, P., & Amorim, M.-A. (1992). Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 785–800.
- Povel, D. J., & Collard, R. (1982). Structural factors in patterned finger tapping. *Acta Psychologica*, *52*, 107–123.
- Rafal, R. D., Inhoff, A. W., Friedman, J. H., & Bernstein, E. (1987). Programming and execution of sequential movements in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *50*, 1267–1273.
- Rand, M. K., Hikosaka, O., Miyachi, S., Lu, X., & Miyashita, K. (1998). Characteristics of a long-term procedural skill in the monkey. *Experimental Brain Research*, *118*, 293–297.
- Rand, M. K., Hikosaka, O., Miyachi, S., Lu, X., Nakamura, K., Kitaguchi, K., et al. (2000). Characteristics of sequential movements during early learning period in monkeys. *Experimental Brain Research*, *131*, 293–304.
- Reason, J. (1979). Actions not as planned: The price of automatization. In G. Underwood & R. Stevens (Eds.), *Aspects of consciousness, Volume 1: Psychological issues* (pp. 67–89). London: Academic Press.
- Reber, P. J., & Squire, L. R. (1998). Encapsulation of implicit and explicit memory in sequence learning. *Journal of Cognitive Neuroscience*, *10*, 248–263.
- Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychological Review*, *77*, 481–495.
- Rhodes, B. J. (2000). *Learning-driven changes in the temporal characteristics of serial movement performance: A model based on cortico-cerebellar cooperation*. Unpublished doctoral dissertation, Department of Cognitive and Neural Systems, Boston University, Boston.
- Rhodes, B. J., & Bullock, D. (2002a). A scalable model of cerebellar adaptive timing and sequencing: The Recurrent Slide and Latch (RSL) model. *Applied Intelligence*, *17*, 35–48.

- Rhodes, B. J., & Bullock, D. (2002b). *Neural dynamics of learning and performance of fixed sequences. I: Data interpretation and model architecture*. Boston University Technical Report CAS/CNS-2002-005.
- Rijntjes, M., Dettmers, C., Buchel, C., Kiebel, S., Frackowiak, R. S. J., & Weiller, C. (1999). A blueprint for movement: Functional and anatomical representations in the human motor system. *Journal of Neuroscience*, *19*, 8043–8048.
- Rosenbaum, D. A. (1987). Successive approximations to a model of human motor programming. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 21, pp. 153–182). San Diego, CA: Academic Press.
- Rosenbaum, D. A., Inhoff, A. W., & Gordon, A. M. (1984). Choosing between movement sequences: A hierarchical editor model. *Journal of Experimental Psychology: General*, *113*, 372–393.
- Rosenbaum, D. A., Kenny, S. B., & Derr, M. A. (1983). Hierarchical control of rapid movement sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 86–102.
- Rosenbaum, D. A., & Saltzman, E. (1984). A motor-program editor. In W. Prinz & A. F. Sanders (Eds.), *Cognition and motor processes* (pp. 51–61). Berlin: Springer-Verlag.
- Rosenbaum, D. A., Saltzman, E., & Kingman, A. (1984). Choosing between movement sequences. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes* (pp. 119–134). Hillsdale, NJ: Erlbaum.
- Sakai, K., Hikosaka, O., Miyachi, S., Takino, R., Sasaki, Y., & Putz, B. (1998). Transition of brain activation from frontal to parietal areas in visuomotor sequence learning. *Journal of Neuroscience*, *18*, 1827–1840.
- Sakai, K., Kitaguchi, K., & Hikosaka, O. (2003). Chunking during visuomotor sequence learning. *Experimental Brain Research*, *152*, 229–242.
- Salamé, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, *21*, 150–164.
- Salthouse, T. A. (1986). Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological Bulletin*, *99*, 303–319.
- Seidler, R. D., Purushotham, A., Kim, S.-G., Ugurbil, K., Willingham, D., & Ashe, J. (2002). Cerebellum activation associated with performance change but not motor learning. *Science*, *296*, 2043–2046.
- Semjen, A., & Garcia-Colera, A. (1986). Planning and timing of finger-tapping sequences with a stressed element. *Journal of Motor Behavior*, *18*, 287–322.
- Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*, 367–447.
- Shima, K., Mushiake, H., Saito, N., & Tanji, J. (1996). Role for cells in the presupplementary motor area in updating motor plans. *Proceedings of the National Academy of Sciences*, *93*, 8694–8698.
- Shima, K., & Tanji, J. (2000). Neuronal activity in the supplementary and presupplementary motor areas for temporal organization of multiple movements. *Journal of Neurophysiology*, *84*, 2148–2160.
- Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117–152). New York: Academic Press (Reprinted in *Perception and production of fluent speech*, pp. 469–505, by Cole, R.A. (Ed.), 1980, Hillsdale, NJ: Erlbaum).
- Sternberg, S., Wright, C. E., Knoll, R. L., & Monsell, S. (1980). Motor programs in rapid speech: Additional evidence. In R. A. Cole (Ed.), *The perception and production of fluent speech* (pp. 507–534). Hillsdale, NJ: Erlbaum.
- Tanji, J., & Shima, K. (1994). Role for supplementary motor area cells in planning several movements ahead. *Nature*, *371*, 413–416.
- Taylor, J. (Ed.). (1932). *Selected writings of John Hughlings Jackson*. London: Hodder and Stoughton.
- Treisman, A. M. (1988). Features and objects: The fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology*, *40A*, 201–237.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.
- Vakil, E., Kahan, S., Huberman, M., & Osimani, A. (2000). Motor and non-motor sequence learning in patients with basal ganglia lesions: The case of serial reaction time (SRT). *Neuropsychologia*, *38*, 1–10.

- van Mier, H., & Hulstijn, W. (1993). The effects of motor complexity and practice on initiation time in writing and drawing. *Acta Psychologica*, *84*, 231–251.
- Verwey, W. B. (1994). Evidence for the development of concurrent processing in a sequential keypressing task. *Acta Psychologica*, *85*, 245–262.
- Verwey, W. B. (1995). A forthcoming keypress can be selected while earlier ones are executed. *Journal of Motor Behavior*, *27*, 275–284.
- Verwey, W. B. (1996). Buffer loading and chunking in sequential keypressing. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 544–562.
- Verwey, W. B. (1999). Evidence for a multi-stage model of practice in a sequential movement task. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1693–1708.
- Verwey, W. B. (2001). Concatenating familiar movement sequence: The versatile cognitive processor. *Acta Psychologica*, *106*, 69–95.
- Verwey, W. B. (2003a). Processing modes and parallel processors in producing familiar keying sequences. *Psychological Research*, *67*, 106–122.
- Verwey, W. B. (2003b). Effect of sequence length on executing familiar keying sequences: Lasting segmentation and preparation? *Journal of Motor Behavior*, *35*, 343–354.
- Verwey, W. B., & Clegg, B. (in press). Effector dependent sequence learning in the serial RT task. *Psychological Research*.
- Verwey, W. B., & Eikelboom, T. (2003). Evidence for lasting sequence segmentation in the discrete sequence production task. *Journal of Motor Behavior*, *35*, 171–181.
- Verwey, W. B., Lammens, R., & van Honk, J. (2002). On the role of the SMA in the discrete sequence production task. A TMS study. *Neuropsychologia*, *40*, 1268–1276.
- Verwey, W. B., & Wright, D. (2004). Effector-independent and effector dependent learning in the discrete sequence production task. *Psychological Research*, *68*, 64–70.
- Vicari, S., Bellucci, S., & Carlesimo, G. A. (2000). Implicit and explicit memory: A functional dissociation in persons with Down syndrome. *Neuropsychologia*, *38*, 240–251.
- Ward, N. (1994). *A connectionist language generator*. Norwood, NJ: Ablex Publishing.
- Wickelgren, W. A. (1969). Context sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, *76*, 1–15.
- Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1047–1060.
- Woodworth, R. S. (1899). The accuracy of voluntary movement. *Psychological Review Monographs*, *3*(Suppl. 2, Whole No. 13), 1–114.
- Wright, C. E. (1990). Generalized motor programs: Reexamining claims of effector independence in writing. In M. Jeannerod (Ed.), *Attention and performance XIII* (pp. 294–320). Hillsdale, NJ: Erlbaum.