MODULARITY OF SEQUENCE LEARNING SYSTEMS IN HUMANS

Steven W. Keele and Tim Curran

Department of Psychology
College of Arts and Sciences
University of Oregon
Eugene, OR 97403
USA

1995

INTRODUCTION

Humans excel at a variety of learned and highly skilled activities in which complex sequential behavior is distributed over time. The major theme of this chapter concerns the hypothesis that sequence learning and production of sequences of activities involves not a single function, but rather is made up of multiple components. For example, in playing a piano, pitch is mapped to key position and key position is mapped to the motor system for bringing the arms, hands, and fingers to the keys. In addition to this spatial mapping, the pianist must learn the sequence of notes or keys that correspond to a piece of music. The sequential representation must indicate not only which note or key is next in a series, but must also specify the intervals at which the keys should be hit and with what intensity. In other activities, dancing for example, trajectory through space, and not just the target of movement, must be specified. It is likely that some of these functions are independent of one another, both in the psychological sense that one function can be affected with minimal or no influence on another, and in a neurobiological sense in that they depend on different brain regions. This chapter will focus on a selected aspect of skill, the representation of learned sequences, and will consider only those representations that specify the succession of events. One of the issues to be addressed is the relationship between the representation of a sequence and the motor system that actually produces the sequence. Evidence will be presented that sequence representation is relatively abstract and independent of the implementation system. A second line of evidence to be presented suggests that the sequential representation itself has constituent parts or modules.

Without a theory to describe the components of sequential representation and performance, it is difficult to design a focused investigation of the neurobiological underpinnings of skill. A considerable amount of research on "procedural learning" has been based on an assumption, not always stated, that sequence learning is less advanced and less differentiated into functions than is verbal or "declarative" learning. Moreover, it is often assumed that sequence learning occurs in some putative motor system of the brain, such as

DEPARTMENT OF THE NAVY
OFFICE OF NAVAL RESEARCH
SEATTLE REGIONAL OFFICE
1107 NE 45TH STREET, SUITE 350
SEATTLE WA 98105-4631

IN REPLY REFER TO:

4330
ONR 247
11 Jul 97

From: Director, Office of Naval Research, Seattle Regional Office, 1107 NE 45th St., Suite 350, Seattle, WA 98105
To: Defense Technical Center, Attn: P. Mawby, 8725 John J. Kingman Rd., Suite 0944, Ft. Belvoir, VA 22060-6218

Subj: RETURNED GRANTEE/CONTRACTOR TECHNICAL REPORTS

1. This confirms our conversations of 27 Feb 97 and 11 Jul 97. Enclosed are a number of technical reports which were returned to our agency for lack of clear distribution availability statement. This confirms that all reports are unclassified and are "APPROVED FOR PUBLIC RELEASE" with no restrictions.

2. Please contact me if you require additional information. My e-mail is *silverr@onr.navy.mil* and my phone is (206) 625-3196.

ROBERT J. SILVERMAN

the basal ganglia or the cerebellum. If sequential behavior involves a complex of modules, however, it seems more likely that different neural systems would provide different components, resulting in distributed representation. Evidence that sequence representation is independent of motor implementation systems suggests that sequence representation originates outside of brain regions that are devoted primarily to selection of particular motor effectors and to actual motor production. Evidence for complex sequential structures and for different modules of representation suggests that a number of different brain regions are involved in sequence representation. A long-term goal of psychological studies is to work out the complexities of sequence learning from a psychophysical point of view in the hope that it will facilitate the analysis of the neural systems that underlie it.

The fact that less research has been done on sequential learning and behavior than on other domains such as verbal memory does not mean that it is more primitive or less important. Speech and language are pinnacles of human achievement. Speech requires the sequencing of a small set of phonemes into a myriad of words, and the sequencing of these words to produce phrases and sentences. Clearly, sequential learning is a prominent aspect of human language.

Besides the sequential aspects of language – be they expressed in speech, writing, sign, or typing – humans also exhibit impressive sequential behavior in other domains. They express music in song, in instrument, in dance. They knit, build cabinets, and acquire the complicated skills of sports. Such widespread capabilities of learning new and exotic forms of sequential behavior indicate that specialized brain systems for sequential learning in humans may generalize beyond the domain of language. Although some theorists have suggested that a key in human evolution is the development of language-specific brain systems, our own investigations have been guided by the notion that humans have evolved mechanisms especially attuned to learning sequential constructions and that subserve both language and nonlanguage.

This idea was articulated in more general form some years ago by Rozin (1976). He suggested that in the course of evolution, particular computational mechanisms arise to solve particular animal problems. In humans, and to a lesser extent in other animals, the computational mechanisms often have evolved further, to the extent that they have become separable from the task of origin and generalizable to other tasks (c.f., Greenfield, 1991, regarding sequential representation in infants, adults, and chimpanzees). This accessibility of a computational module by a variety of inputs and outputs, Rozin argued, lies at the heart of human intelligence. One of Rozin's primary examples concerns phonetic representation. Part of human speech capability stems from decomposition of speech sounds into elementary phonemes that can be reordered to produce different words. In humans, the phonetic representation that subserves speech can also be tapped into by a uniquely human invention, visual symbols called graphemes that can map onto phonemes and serve as a basis for reading. Thus, a module involved in speech is accessible through vision, a different input than anticipated in the course of evolution. A similar view of modularity has been advanced from a neurobiological perspective by Mesulam (1985, 1990) who suggests that local neural networks underlie specific cognitive operations. These local networks participate in a variety of complex behaviors through their large-scale interaction with other computational networks.

Recently, other examples of common computational modules that underlie diverse human activities have been described. One such computation that has inspired much of our sequence work concerns timing. Ivry (e.g., Ivry and Keele, 1989; Ivry and Gopal, 1992; Keele and Ivry, 1991) has presented evidence that a timing mechanism, operating in the range of a few hundred milliseconds to a second or two and localized in the cerebellum, underlies a variety of motor and perceptual tasks. Evidence for this idea comes from the following observations: (1) Timing of intervals in repetitive motor tapping is disrupted by lesions of the lateral cerebellum; (2) Speech dysarthria resulting from cerebellar damage reflects disruption of precise temporal relationships between speech components, as in the voice onset time of

stop consonants, but does not affect nontemporal speech properties such as vowel formant structure; (3) Perceptual judgments of time between auditory events are disrupted by cerebellar damage, but loudness judgments of the same events are not; (4) Judgments of the velocity of moving visual displays, which depend on temporal information, are impaired in patients with cerebellar damage, but positional judgments of the same displays are not.

It has been argued that the same lateral regions of the cerebellum are necessary for classical conditioning (e.g., Thompson, 1986; 1990). Many forms of conditioning involve very precise timing in which the interval between a conditioned stimulus and a conditioned response corresponds to the interval between the conditioned stimulus and an unconditioned stimulus. This timing relationship often has adaptive value. For example,a conditioned eye blink that temporally anticipates a noxious stimulus to the eye may prevent the noxious stimulus from having damaging effects. Lesions to the cerebellum impair or abolish precisely-timed forms of classical conditioning, but have little effect on other types of conditioning, such as emotional conditioning. For example, in experiments where a tone is followed by a small electrical shock near the eye, lesions of the cerebellum may affect the linkage of the tone to eyeblink, but they do not affect the linkage of the tone to the autonomic response of change in heart rate elicited by the same shock (Lavond et al., 1984). Moreover, there is evidence that different systems within the cerebellum play different roles. Lesions of the nucleus interpositus of the cerebellum abolish precisely timed conditioned responses, while lesions of cerebellar cortex leave conditioning intact but with responses occurring at inappropriate times (Perret, Ruiz, and Mauk, 1993).

These studies suggest that in humans a particular class of computation, timing in the millisecond range, is separable from the performance of an individual task, and that the cerebellar cortex plays an essential role in the timing computation. We call a system that performs a class of computations and that can be interfaced with different inputs or outputs a module.

In this chapter we examine other components that contribute to skill, concentrating on psychophysical studies of sequence learning. We provide evidence that sequence representation is modular in the sense that it is separable from the motor systems that actually implement movement. Thus, sequencing resembles timing in that an abstract relationship is transferrable among different input/output systems. Secondly, we provide evidence for different sequential learning systems that are in certain respects independent of one another. We review some network models of sequence learning that are beginning to provide insight into possible computational mechanisms of learning. In addition, we discuss ways in which the psychophysical studies could be applied to an analysis of neural mechanisms involved in sequencing.

A MODEL TASK FOR STUDYING SEQUENTIAL REPRESENTATION

In biology and psychology, it is common to study particular, species-specific behaviors. To study sequencing, one might examine behaviors as diverse as language and speech, locomotion, musical performance, birdsong, or mouse grooming. However, the fact that humans are adept at learning a variety of sequences, many of which probably depend on common computational systems, has motivated the design of model tasks that differ from most naturalistic tasks, but have certain experimental advantages. A model task should comprise critical features of important human sequential tasks but use simple procedures amenable to experimental manipulation. The model task should be learnable within a short time frame. Appropriate model tasks can also be employed with animal and infant subjects in order to relate psychologically defined components to neural substrates and their development.

Several different model tasks have been developed to study sequencing. In one pioneering effort, Restle and Burnside (1972) used a linear array of six lights that corresponded to six response buttons. The lights came on successively in patterns such as 1234666662323543, where the numbers refer to lights from left to right. A subject's task was to learn to press a key corresponding to the next anticipated light in a sequence. The lights were presented at a fast pace so that subjects frequently made late responses, responding to one light after a subsequent one had already appeared. Late responses predominated at particular places in the sequence – at the end of a regularly changing sequence in one direction (1234), a sequence in the reverse direction (543), a set of repetitions (66666), or a trill (2323). These break points, where a subject was slow in anticipating the next light, suggested a simple but powerful principle, namely that the internal representation of a sequence had a hierarchic structure. That is, a sequence is stored and retrieved as a series of chunks, each chunk having its own internal structure.

The idea of hierarchic representation was subsequently elaborated by others. Povel and Collard (1982), rather than presenting lights, simply showed subjects a series of numbers (e.g., 321234) that represented the order in which subjects were to press 4 keys in a repeating sequence. The lengths of intervals between successive responses suggested that different individuals parsed such sequences in different ways. For example, some subjects might parse the sequence as a backward run (321) followed by a forward run (234), exhibiting a rather large transition time between 1 and 2. Other subjects might exhibit a parsing such as (32) (1234). Yet others might represent within the sequence the trill 212 preceded by 3 and followed by 34. The important point is that one and the same sequence is subject to different internal and hierarchic representations that can be deduced from the temporal output structure. Very similar ideas have been suggested by Rosenbaum (1987).

Not all investigations have used key pressing as a model task for sequencing. Gordon and Meyer (1989) taught subjects short sequences of 4 nonsense syllables. Sometimes they asked subjects to prepare to produce one sequence but then unexpectedly signaled them to perform a different one composed of the same nonsense syllables but differently ordered. Examining the time to reprogram a sequence led them to the now familiar conclusion that the internal representation of an event sequence, rather than being a linear string of associations, actually was hierarchic. In these experiments, a string of four elements was coded as two concatenated strings of two elements each.

These studies all involved explicit learning, meaning that subjects were either told or otherwise became aware of the exact order in which events occurred. In other situations, where events occur and are responded to in some particular order, the sequential structure is not apparent to the learner. This type of learning is referred to as implicit learning. Although a performance criterion may indicate that the sequence has been learned, the subject is not aware that any learning has taken place. For example, when children learn language prior to beginning school, they typically are not told the rules of word ordering that constitute the grammar of their language. Nevertheless, they are capable of producing correct sequences. It is not uncommon even for adults to be unable to describe the rules that govern their choice of word order, even though their grammar is invariably correct. For review and discussion of the distinction between implicit and explicit learning see Berry (1994), Reber (1989), and Shanks and St. John (1994).

For investigating questions involving explicit versus implicit sequence learning, a paradigm originally developed by Nissen and Bullemer (1987) has proven useful. Subjects view a screen with 3, 4, or 5 designated positions in a horizontal line. On each trial a visual signal, such as an X-mark, can appear at any position. Beneath the screen are corresponding response keys. The subject's task is to press the key that corresponds to the position of the visual signal – key 1 for signal position 1, etc. Reaction time is measured from signal onset to key press. Following a key press, and usually after a fixed interval (e.g. 200 ms), the next

signal occurs. Typically blocks of about 100 signals are presented, after which there is a short rest period.

Within a block of trials, the signals can occur in either random or sequentially structured orders. Random signals are usually presented with the constraint that the same signal is not presented twice in succession. Sequential signals occur in specific orders. In a sequence designated 13232....., signals occur at three positions, numbered 1 through 3 from left to right. A set of 5 signals occurs in the order designated, after which, without any discernible break, the sequence recycles. The first signal on a block of trials can start at any particular position within the sequence. The subjects' task is simply to respond to each signal as it occurs, trying to respond as rapidly as possible. Subjects learn the sequence structure, whether they report awareness or not. Sequence learning can be quantified by comparing subjects' reaction times when events occur in sequence with their reaction times when events occur at random. Such an index provides a measure of performance learning that does not require awareness of the learning.

Nissen and Bullemer (1987) introduced another manipulation to examine the role of attention in sequence learning. In the typical experiment, there is a 200 ms interval between one response on the primary task and the presentation of the next visual signal. During that interval a high or low pitched tone can be inserted, and subjects are asked to count the high-pitched tones. Usually performance on this distraction task itself is not of great interest; rather the distraction is used to interfere with subjects' attending to the relationship between successive events of the primary task.

The important contribution of Nissen and Bullemer's paradigm is that it allows assessment of sequential learning for sequences of a variety of types under conditions where explicit instruction is provided, or where no information about the sequence of stimulus is given. In addition, these experiments can be performed under conditions of distraction or full attention.

INDEPENDENCE OF SEQUENTIAL REPRESENTATION FROM THE MOTOR SYSTEM OF EXECUTION

The modular theory of sequence processing suggests that the same internal representation of a sequence of events can be interfaced with diverse motor systems for executing the sequence. Rozin (1976) had developed this general argument based on evidence that phonetic representations underlie not only speech production and perception but also the reading of visually presented words. A number of lines of evidence are consistent with the view. At the informal level, it is often noted that writing style is remarkably similar for the same person, whether it is performed on a small scale by the hand or on a large scale by the arm. Even exotic effectors such as head movements or elbow movements produce similar writing styles (Bernstein, (1947) as reported in Keele, Cohen and Ivry, 1990; Raibert, 1977). Wright (1990) averaged multiple writing samples to eliminate sample-by-sample variation and noted even more remarkable similarity between hand writing and arm writing, though some effector differences emerged as well.

These informal observations suggest that the same internal description of space guides different effectors in the production of figures, consistent with a view put forward by Berkenblit and Feldman (1988): "There is a neuronal level that creates an abstract image (verbal or graphic) of the forthcoming movement (a circle, line, etc.). Then a combination of effectors and a coordinative structure is specified...."

The informal observations of letter similarity across effector systems together with the notion of an abstract image suggest that at least certain aspects of the representation of a motor act involved in drawing a single graphic figure, such as a letter or a geometric shape, are accessible by different effectors. However, these observations raise the question of how

low in a hierarchy of motor acts such modularity descends, and at what level motor representation becomes specifically designed for the responding effector.

In a study reported in preliminary form, Wright and Lindemann (1993) had subjects practice writing particular letters with the *nondominant* hand. As would be expected, fluency in producing the letters improved with practice, but the important question was how the improved fluency transferred to nonpracticed letters when writing with the nondominant hand. Nonpracticed letters that shared the same strokes as the practiced letters, even though the strokes were arranged differently, were produced as well as the letters that had been practiced. However, letters composed of strokes not contained in the practiced letters were not executed as well. These results suggest that in handwriting, practice with a specific effector is confined to the level of strokes, i.e., skill improvement is effector dependent. In contrast, the mechanisms involved in arranging strokes into letters must be effector independent because only the dominant hand would have had extensive experience with the stroke arrangements in the nonpracticed letters. Thus, practice using a specific effector improves stroke production but such practice is not necessary for stroke assembly, implying that some basis for the assembly already exists.

Given that specifications of motor action above the stroke level for single letters can be shared at least across hands, one would expect that even higher levels of description such as the specification of a series of letters in a written word can also be shared among different motor systems. Hillis and Caramazza (1988) examined two patients with posterior cortical damage both of whom suffered partial unilateral neglect. The patient with right hemisphere damage tended to make handwritten spelling errors on the left edge of words, often misordering letters, as in writing "rpiest" instead of "priest". The patient with left hemisphere damage tended to make errors on the right edge of the words, again often misordering letters. Of particular interest is the observation that when the patients were asked to spell words orally, both patients made a similar proportion and type of spelling errors as in handwriting. Oral spelling occurs over time, not space, suggesting that the common "locations" of spelling errors for oral and written spelling both made use of a common internal specification of letter order.

In summary, the specification of serial order appears to be abstract in the sense that different effector systems can draw upon the same internal description of movement through space.

The view that serial specification of letters or phonemes is completely independent of the effector system that will execute the movements is not universally shared. Although most studies of the relationship between sequential specification and the motor system imply some degree of independence, they do not necessarily indicate that there is complete independence. A prominent theory to describe sequential motor behavior developed by Jordan (1986; 1993) builds sequential representation into a network that contains effector-specific components, rather than separating sequential specification from the motor systems of action. His model can perhaps best be appreciated by considering the task of ordering phonemes to produce words in speech.

In Jordan's model network, two types of input units are combined via hidden units to jointly specify the features of the next phonemic output (this model is more fully described in a later section, c.f. Figure 8). One set of inputs is a so-called plan, which can be thought of as a global representation of the word. The other input is a set of state units that essentially maintain decaying memories of the phonemes already emitted in a sequence. At the beginning of a phonemic sequence, the state units are all initialized to zero, so the first phoneme is determined exclusively by the plan units. After the first phonemic output, some of the state units change, reflecting the speech features that were involved in the immediately preceding phoneme. As a result the next output is a product of the original plan, plus the new state. The output units in the model are not abstract representations of phonemes that could be fed to motor speech apparatus. Rather the outputs are features of speech, such as lip rounding, tongue position, etc., and the state units that preserve information about prior outputs do not

represent memories of abstract phonemes but instead represent the speech features that made up the phoneme. As a result of learning, the network develops the ability to produce the next appropriate set of speech features upon receipt of feedback that the speech features of the preceding phoneme have been emitted. This network, therefore, produces sequential behavior, but it does so within a system that has intrinsic speech-related outputs.

What would be the advantage of using speech features as output rather than abstract representations of phonemes that could then be shipped off to another modular network that would produce appropriate motor activity? The reason is that with practice, motor action becomes fluent so that one aspect of movement melds smoothly into another. One goal in constructing a series of movements is to minimize the amount of change in a motor effector during the transition from one movement to another. The mutual interaction of nearby movements is called co-articulation. Co-articulation maximizes the efficiency of movement with the sole constraint that the individual components still remain interpretable. To produce a smooth series of movements presumably requires specification of the movement apparatus. Thus, in typing, a particular kind of combined finger movement would be required to produce two keystrokes in quick succession. In speech, the motor apparatus might need a very different kind of motor control to make the transition between adjacent phonemes fluent. In actual fact, research in both speech and typing shows a tremendous amount of co-articulation between successive motor actions. Although some co-articulation might be explained in other ways, it was this feature that motivated Jordan's decision to incorporate the executing motor organ into the actual sequential representation.

Thus, some evidence suggests that sequential specification of motor acts is independent of the selection of motor effectors that will produce the actions. Nevertheless, there is also at least a theoretical basis for questioning complete modular separation.

Our own studies have used quantitative assessments of the transfer of learning to distinguish between a theory of modular representation and one in which sequential specification is intrinsic to the effector system. All of our studies have used a variant of the design developed by Nissen and Bullemer (1987).

To determine whether different motor effectors share a common sequence representation (Keele et al., 1995), subjects were trained to respond to sequential stimuli using one of two motor systems. The first involved using three fingers (index, middle, ring) to depress three keys and the second involved moving the arm back and forth to strike the keys with the index finger only.

Subjects began with two blocks of visual signals presented at random. The visual signals were an X-mark that appeared at various locations. The first experiment was conducted without a distraction task. Subjects were told to respond as rapidly as possible to each visual signal by pressing a corresponding key. Although they were told nothing about the presence of a sequence, many became aware of its presence. Different subjects received different sequences, but all were five elements in length, involving 3 positions, 2 of which were repeated within the sequence. An example is 13232...., where the numbers refer to order of positions on a screen at which a visual X-mark appeared. A preliminary *practice* period with successive events presented at random familiarized the subjects with the mapping from stimuli to key-press responses, but did not allow any learning of sequential order. The random practice period also obscured the fact that later events might be presented in sequence. Following practice, subjects typically received 6 to 8 *learning* blocks in which signals were presented in sequence. Some subjects were not told about the sequence. Following learning, subjects entered a *transfer* phase in which some were required to switch to a different effector system. At transfer, half the subjects continued to use the same effector system as in initial learning; the other half changed to the previously unpracticed motor system. During transfer, subjects typically received one block of random trials to familiarize them with the new response arrangements, followed by random, then sequenced, then random blocks. The difference in reaction time between random and sequenced blocks during this final phase was taken as a

measure of the amount of sequential information acquired during learning that transferred to new conditions with a different response mode. If the sequential representation were independent of the motor system of execution, there should be complete transfer of knowledge.

Figure 1 shows the results. For the group that used the same effector during the transfer phase as during initial learning, the reaction time for events that occurred in sequence was about 130 ms less than for random series of stimuli. For subjects who switched to a previously unpracticed effector, the reduction in reaction time was approximately the same and did not differ statistically from the former, suggesting a common sequence representation for the two effector systems.

However, it is possible that the random blocks preceding the sequence test extinguished learning, and the 130 ms sequence advantage actually represents new learning on the single sequence block of the transfer phase. To examine that possibility, a control group of subjects received random events throughout the entire "learning" period and changed effector system during the transfer period, so that the sequence was introduced on a single test block. Those subjects also showed faster reaction times on the sequence block than on the surrounding random blocks, suggesting that some learning occurred on that block alone. Nevertheless, the sequence advantage was substantially and reliably less than that for subjects with previous practice on the sequence.

These results suggest that *all* sequence knowledge acquired in the context of responding with one effector system transfers to a different effector system, because amount of transfer did not depend on whether or not the effector changed. This finding further suggests that sequential representation resides in a separate module from implementation systems.

Figure 1. Performance of two groups of subjects who were presented with sequences during the learning phase. During transfer, one group changed the responding effector from fingers to arm or vice versa. The other retained the same effector. Sequence learning is indicated by the difference between sequence block 13 and random blocks 12 and 14, as shown in the inset. The letter R on the abscissa indicates a random block and the letter S a sequence block except for a control group that received random events on all of blocks 1-12 and 14. All blocks are under single-task conditions.

A hallmark of explicit memory is the flexibility with which it can be expressed (e.g., Cohen and Eichenbaum, 1993). To test whether transfer of sequential knowledge between effectors is due to explicit memory, or whether transfer also occurs when subjects are unaware of the sequence, we used a distraction procedure that has been shown to abolish awareness of a sequence for almost all subjects (Cohen, Ivry, and Keele, 1990; Nissen and Bullemer, 1987). A tone was inserted in the short 200 ms interval between each keypress response and the next visual stimulus of the primary task. At the end of a block of trials, subjects reported the number of high-pitched tones.

Figure 2 shows the results of this experiment. The addition of the secondary task caused reaction times on the primary task to increase, and it also reduced the total amount of sequence learning as assessed by the difference between random and sequence conditions during the transfer period. Nonetheless, sequence learning was reliable, and the amount of learning did not depend on the effector that was used during transfer. A control group that had not experienced the sequence until the critical test phase showed no evidence of learning. Thus, under conditions of distraction, the shorter reaction time for the sequence is the exclusive result of learning prior to transfer rather than new learning during the transfer phase. This experiment suggests that sequential knowledge is independent of the motor system that expresses it, even under distracting conditions that minimize explicit learning.

Figure 2. Identical experiments as in Figure 1, except that all blocks are under dual-task conditions, in which tones were presented between visual stimuli and subjects were instructed to report the number of high-frequency tones.

The two experiments just described suggest that sequential representation is not in the motor system because the knowledge is completely transferable from one motor effector to another. However, they do not distinguish between two other possibilities having to do with the nature of the representation. The form of the sequence representation could be either the order in which the stimuli occur, in this case a visual/spatial representation, or the order in which keys are pressed. Presumably, key press order is not the same as a motor code, because the same response key can be pressed with different motor effectors.

To differentiate between these two types of sequence representation, a third experiment was conducted. It introduced verbal responses in addition to finger responses so that transfer could occur with changes in the nature of the response. Like the second experiment, this one

employed a distraction task throughout. During learning, one group of subjects responded verbally to signal position with the words left, middle, and right; the other group responded manually with key presses. During transfer, both groups responded verbally. Verbal reaction times were measured from voice-onset times.

The results shown in Figure 3 in some ways resemble those of the first two experiments, but differ in other ways. Some sequence knowledge acquired manually did transfer to verbal responding. That sequence knowledge, though small in magnitude, was statistically greater than that exhibited by a control group making manual responses that had only been presented with random events during training. These results suggest that, at least some of the sequence knowledge acquired in responding to a series of visual events describes the order of the events, not the order of responding. It is also the case, however, that the group who practiced with verbal responses throughout learning exhibited greater sequential knowledge during the transfer phase than the group that had practiced manually and transferred to verbal responding.

Figure 3. Performance of two groups of subjects, one group that practiced with manual responses, and a second that practiced with verbal responses. A control group also practiced with manual responses, but events were random. During transfer, all groups used verbal responses. All blocks are under dual-task conditions.

There are at least two possible explanations for incomplete transfer of sequence learning between manual and verbal responding. Less sequence learning might occur with manual responses than with verbal responses, so there would be less information to transfer from the manual to the verbal mode. Second, although some sequence information resides in stimulus order, some may also reside in response order. Further experiments will be necessary to distinguish among these possibilities.

Other researchers have also provided evidence that sequential information resides in a code independent of the motor system. In a design somewhat similar to our own, Stadler (1989) found that sequential information about perceptual events transferred from key pressing with one set of fingers to responding with a different set of fingers. Mayr (1995) also demonstrated that some sequential representation is tied to stimulus representation,

completely independent of response requirements. Subjects in Mayr's study pressed keys that corresponded to the identity of geometric shapes. The shape on a given presentation appeared in one of four different locations, but location was irrelevant to response selection. Unknown to the subjects, the geometric shapes and hence the order of responses occurred in one particular sequential order. The locations of the shapes occurred in a different and uncorrelated order. By occasionally reverting to random order either in shapes or in positions and observing declines in reaction time, Mayr was able to show that subjects had acquired sequential knowledge not only of the upcoming shape, which determined the response, but also of its position, despite the fact that position was not the response determinant subjects had been instructed to use. When learning took place without distraction several subjects became aware of the sequences, but similar results occurred in a tone-distracted study that virtually blocked awareness.

A study by MacKay (1982), using German-English bilinguals to examine sequence transfer, provides additional insight into the nature of the sequence representation. He presented subjects with sentences in one language or the other and observed their improvement in speaking speed over 12 repetitions of the same sentence. When the sentence was then read aloud in the other language, preserving the same ordering of concepts, MacKay found 100 percent transfer of the previous sequence learning. While most people would intuitively predict some transfer, the important observation is that transfer was complete. This outcome indicates that the speed improvements were localized to a sequential representation not only more abstract than particular movements of speech apparatus but also more abstract than a specific language or word order. In another of MacKay's experiments, subjects spoke a randomly ordered series of words, repeating them in the same random order each time. When switched to the alternate language, no transfer of learning occurred.

How might MacKay's apparently discrepant findings for the two situations be explained? He proposed a theory of hierarchic representation of sequences (see MacKay, 1987 for a much expanded treatment of his theory and many other sequential phenomena). His model divides the representational system into three modules – conceptual, language-specific, and motor. High-level sequential representation resides in the conceptual module. A sentence is represented first at the level of abstract object-action concepts. At successively lower levels of a hierarchy, the complex concepts are differentiated until they correspond to individual concepts that can be denoted by different words in a different language. It is important to note, however, that even at this level, concepts are not words; the same concept can underlie both an English and a German word. For accomplished bilinguals, who already have fluent articulatory abilities, the novelty of a new sentence is restricted primarily to the highest levels. Thus, when a novel sentence is practiced, learning is restricted primarily to new conceptual structure. It is this information that transfers from one language to the other. The reason why transfer fails for randomly ordered words as opposed to those words embedded in a sentence, is that conceptual structure helps convey the meaning of individual words, and such structure is lacking for random words. Consider a word such as "right" that has numerous meanings, referring variously to spatial position, lack of error, political orientation, etc. Sentence structure helps restrict the conceptual meaning. Strictly serial order representation with no hierarchical structure fails to specify precise meaning and hence limits transfer to words of another language.

Once the conceptual representation reaches its terminal level, individual concepts are translated into language-specific words through a second module. This module too contains hierarchic representation. A word is successively broken down into components, first into syllables, and ultimately into a phonetic representation. The phonetic representation is then interfaced with a third module that specifies articulatory components.

MacKay's theory and his own empirical observations are consistent with the present evidence that sequential representation of action is separable from the motor system that implements the action. In MacKay's theory, articulatory control systems can produce

individual articulations, but do not contain any information about what subsequent articulations will be. Any anticipation of articulation comes not from prior articulation per se but from higher levels of hierarchical representation that are functionally separable from articulation.

In summary, a number of studies, all using a transfer methodology, have suggested that sequential representations that guide a sequence of actions reside in a module that is separable from the effector system itself. Although these studies suggest that separability of sequential representation from implementation applies to both conditions of attended and unattended learning, there is evidence that sequential memories acquired under conditions of distraction involve at least partially separate modules from those requiring attention.

INDEPENDENT ATTENTIONAL AND NONATTENTIONAL SEQUENCE REPRESENTATIONS

A number of experiments have sought to specify sub-components of the sequence learning system (Nissen and Bullemer, 1987; Nissen, Willingham, and Hartman, 1989; Nissen, Knopman, and Schacter, 1987). These studies have used relatively complex sequences of the type, in which visual signals are presented at successive positions, with key-pressing reaction time used as a measure of sequence learning. Nissen and colleagues found that a distracting task prevented subjects from learning complex sequences that were easily learned without distraction. As long as there was no distraction, even patients with Korsakoff's syndrome, who suffer severe amnesia as a result of chronic alcoholism, were able to learn the sequence, although they were unable to express awareness of it (Nissen and Bullemer, 1987; Nissen, Willingham, and Hartman, 1989). Administration of scopolamine, an anticholinergic drug, reduced sequence awareness in normal subjects but did not prevent sequence learning (Nissen, Knopman, and Schacter, 1987). Thus, the work of Nissen's group suggests that attention, in terms of freedom from distraction, is needed for learning a complex sequence, but that attentional learning does not necessarily lead to awareness (see also Willingham, Nissen, and Bullemer, 1989).

In our laboratory, we (Cohen, Ivry, and Keele, 1990) found that under certain circumstances, even with a concurrent secondary task, certain types of sequences were learned. Sequences of the sort 13232.... where numbers refer to spatial positions, were learned, as were slightly longer sequences involving more elements, of the sort 132314..... However, when sequences of the type 132312.... were used, a distraction task blocked sequence learning. The latter two sequences are extremely similar, involving a different event only at one sequence position. Why is one learnable under distraction and the other less so?

We noted that a sequence such as 132312.... has a certain ambiguity. Each possible event occurred twice, but on each occasion, it was followed by a different event. Such a sequence is impossible to learn based only on direct, pairwise associations. The sequence used in all of Nissen's studies has this ambiguous character. We hypothesized that such ambiguity is solvable by a coding mechanism that parses a sequence into chunks, allowing the learning of order within a chunk. This mechanism is essentially one of hierarchic coding. Sequences of the sort 13232... and 132314... have, in contrast, one uniquely occurring event, and more than one unambiguous ordering within a sequence. Thus, in 132314..., event 4 is followed only by event 1 and event 2 is followed only by event 3. Such unique associations, we hypothesized, allowed learning of the entire sequence by a non-hierarchic mechanism. In short, our initial studies led us to postulate two distinct forms of sequence learning, one hierarchic and the other non-hierarchic.

The finding by Cohen et al. (1990) of an interaction between attention and sequence structure led to two extended lines of investigation. The first line (Curran and Keele, 1993) followed up the suggestion that there might be two independent forms of sequence learning,

one of which requires attention but is capable of learning sequences with ambiguous associations, presumably by mechanisms of hierarchic representation. The second line (Keele and Jennings, 1992) involved computational investigations of possible mechanisms of hierarchic and nonhierarchic forms of sequence learning.

We (Curran and Keele, 1993) performed four experiments to examine the hypothesis that there are two independent forms of sequence learning, one requiring freedom from distraction and the other not. For convenience, the two forms of learning will be called attentional and nonattentional, to distinguish between their relative susceptibility to distraction. These two forms of learning, it was hypothesized, do not communicate their contents to one another. When attention is available, the attentional system acquires information in parallel with the nonattentional system. It was hypothesized that the attentional system needs attention not only for acquisition of sequence knowledge, but also for the conversion of that knowledge, once acquired, into either performance or awareness. Finally, we hypothesized that the two learning forms differ in their capability. Specifically, only the attentional system is capable of learning sequences thought to require hierarchic coding; both systems can learn sequences that do not have ambiguity of association.

A first experiment varied the amount of learning that occurs in a nondistracted state by either explicitly telling subjects of the presence of a sequence or not telling them. This manipulation influences the amount of attentional learning and allows a test of the hypothesis that there will be no change in the amount of knowledge in the nonattentional system, because of the assumption that attentional learning is not available to the nonattentional system. Knowledge in the nonattentional system can be assessed by adding a distraction task following initial learning.

There were two basic conditions during training, and two groups of subjects. Under both conditions, the first two blocks of trials were random. Then, the informed learning group was told that signals would occur in a particular order. The order was described, and subjects were given a minute to study it. Different subjects received different signal orders, but all involved four events, two of which occurred twice and two that occurred once in a repeating 6-event cycle (e.g., 143132...). These sequences are learnable with distraction (Cohen et al, 1990).

These same orderings were given to a second set of subjects, but they were not informed that a sequence was present. A questionnaire administered prior to the transfer phase indicated that some of the uninformed subjects had become aware of the sequence on their own and were able to describe parts of it; other subjects did not express awareness. Although the specific awareness criterion used affects the number of subjects placed in one group or the other, the exact dividing point is not critical. The main point is that a group with explicit knowledge, a group that discovered the sequence themselves (more aware), and a group that expressed little or no awareness (less aware) differed in the amount of sequence knowledge exhibited when there was no distraction task. The results shown in Figure 4 were in accord with expectations. When there was no secondary task, variations in degree of explicit knowledge had a clear effect on performance. The informed learning group had very fast reaction times when the sequence was present, and slowed considerably when events returned to random order. Subjects who became aware of the sequence on their own performed as well as informed subjects after some practice. Subjects who expressed little or no awareness of the sequence, still learned under single-task conditions, but showed a somewhat reduced sequence effect.

After a training period without distraction, all groups were transferred to a situation where the tone-counting distraction task was added. Sequence learning was reassessed by comparing performance on blocks of random events and blocks of sequenced events. Despite variations among groups in single-task learning, once the distraction task was added, the three groups performed comparably. All showed significant evidence of residual sequence knowledge, but statistically speaking, the residual knowledge was equivalent for all three

groups. Results of our earlier reported experiments, illustrated in Figures 1-3, indicate that the small sequence effect that occurs under the dual-task conditions was not due to new learning on the single block of sequenced trials during the transfer phase.

The results in Figure 4 are consistent with the hypothesis that variations in attention-based learning are not transferable to the nonattentional system. Such results are rather remarkable. One group of subjects had been told precisely the nature of the sequence, and they could parlay that knowledge into extremely fast performance when there was no distraction. Indeed, mean reaction times of about 200 ms after some practice suggest that at least some of the responses actually anticipated the next stimulus, because in the absence of anticipation, reactions times are seldom that fast. Nonetheless, that knowledge was of no use when a distraction task was added, because performance dropped to that of a group that expressed no awareness of the sequence. A related, and even more powerful point is made in the next experiment.

Figure 4. During single-task learning, one group of subjects was informed of the sequence, another became aware of the sequence on its own, and a third group did not become aware. The figure also indicates performance when a secondary task was added. Random blocks are designated by R and Sequence blocks by S.

In the experiment just described, sequence knowledge was acquired under a condition with no distraction task. A stronger prediction is that knowledge acquired by the non-attentional system under conditions of distraction is equivalent to knowledge acquired when free from distraction. To test this, two groups of subjects were run much as before, except that no diagnostic of sequence knowledge was given during the learning phase. Following two initial dual-task practice blocks with random events, the distraction task was removed for one group, and that group was explicitly told the nature of the sequence. For the other group, not only were they not told about the sequence, but all training occurred under dual-task

conditions. Cohen, Ivry, and Keele (1990) had shown that under such distraction conditions, few if any subjects became aware of the presence of a sequence.

As seen in Figure 5, performance during the transfer phase was poorer in general for the group that had practiced under single-task conditions. Undoubtedly this was because they were less proficient at interweaving the two tasks, given that they had less practice in doing so. Nevertheless, while both groups showed a reliable difference between the sequence test block and the two surrounding random blocks, indicating that they had learned the sequence, that measure was not reliably different between the two groups. The group that had practiced under dual-task conditions and the group that had practiced under single-task conditions showed equivalent transfer of knowledge to the test phase. As before, it seems quite remarkable that a group with explicit knowledge of the sequence showed no better sequence performance than a group without such knowledge, once a secondary task was added. The results suggest that none of the explicit knowledge had transferred to the nonattentional system. In that sense, the two systems are independent.

Figure 5. One group of subjects learned the sequence with no distraction following explicit instruction on the nature of the sequence. The other group learned implicitly with distraction , i.e., was never told about the sequence. The distraction prevented awareness of the sequence.

One potential criticism of these two experiments is that the amount of sequence knowledge during the critical test phase may in fact differ under different conditions, but something about the procedure itself prevents its manifestation. Perhaps the dual-task setting puts some kind of ceiling on the amount of learning exhibited. Thus an experiment was designed to address this concern. If the initial sequence learning occurs under dual-task conditions, presumably learning occurs only in the nonattentional system. If the secondary task were removed following dual-task learning, reaction times should improve. Because there would initially be no knowledge in the attentional system, the benefit of sequential

conditions over random conditions should remain unaltered. It is useful to re-examine Figure 4 in which transfer was in the opposite direction, from single- to dual-task. There, even the unaware group showed a larger sequence effect during single-task conditions than under dual. We take this to mean that as long as there is no distraction, subjects learn more in single-task conditions than in dual. Lack of distraction allows some learning of the type we call "attentional" despite being unaware. This idea is supported by the finding that "attention-based" sequence learning can occur under administration of scopolamine and in patients with Korsakoff's syndrome, both groups with reduced awareness (Nissen, et. al., 1987, 1989). The prediction is, however, that when conditions are reversed, going from dual- to single-task conditions, the sequence effect will be equivalent under both dual- and single-task settings.

To test this prediction, a single group was examined. This group was initially trained under dual-task conditions with one random block inserted to allow assessment of the amount of learning. Then in the transfer phase, the distraction task was removed and sequential knowledge again assessed. The sequence block in the last phase was the sole occasion on which a sequence had been experienced under single-task conditions, and the prediction is that single-task performance would be no better than dual-task performance. The results are shown in Figure 6.

Figure 6. Performance of subjects trained under dual-task conditions and then transferred to single-task conditions.

Reaction times on the first two blocks of single-task conditions were no faster than the preceding dual-task conditions because the last block under dual-task conditions had been with a sequence, while the first two single-task blocks were random. These factors counterbalance each other. The critical point is that there was no statistical difference in performance on the sequence between single-task and dual-task conditions. It appears that during the original dual-task learning, only the hypothesized nonattentional system was available for learning, and only that knowledge source had any useful information about the sequence during single-task transfer.

The experiments described so far show that a nonattentional system can learn sequences that contain some points at which one event uniquely predicts the next event in a sequence. The work of Cohen, Ivry, and Keele (1990) had shown that ambiguous sequences of the sort 132312.... are difficult to learn with distraction. What remains to be demonstrated is that knowledge of sequences of this latter sort, when learned under distraction-free conditions, presumably by the attentional system alone, is blocked when distraction is subsequently added. Such a demonstration would argue that attention is needed not only for learning ambiguous sequences, but also for performance once learning has occurred.

To test this an experiment was designed in which, a single group of subjects initially learned an *ambiguous* sequence under single-task conditions. Because earlier experiments (e.g., Keele and Jennings, 1992) had suggested that ambiguous sequences (e.g., 132312...) sometimes could be learned to a marginal extent even under dual-task conditions, we made the sequence more complicated. We continued to use 3 events but embedded them in a 9-element cyclic sequence such as 132312123.... Following initial single-task training, including the diagnostic test of sequence knowledge, subjects transferred to the dual-task condition and sequence knowledge was again assessed. Subjects showed clear evidence of sequence learning of the ambiguous sequences in the absence of distraction, but subsequent addition of distraction abolished signs of learning. These results suggest that attention is needed not only to code events by place in a sequence, but also to keep track of place in the sequence during performance.

Figure 7. Performance of subjects trained under single-task condition with *ambiguous* sequences, in which each sequence event is followed by a different event depending on the place in the sequence. The figure also shows performance of the same subjects when a secondary task is added.

An observation related to this last experiment was described by Nissen and Bullemer (1987). They note that when subjects were presented with an ambiguous sequence under dual-task conditions, no learning was manifest. Moreover, when the distraction was subsequently removed, not only was there no immediate evidence of sequence knowledge, but subsequent single-task learning showed no acceleration. These observations, when coupled with our own, suggest that sequences entirely composed of ambiguous associations are not stored in a nonattentional memory system.

Recent unpublished work (Goschke, personal communication), has largely supported this hypothesis. Goschke examined sequences of 12 elements, constructed from 6 different signal locations. Each location occurred twice in a sequence, but was in each case followed by a different event. In this paradigm, if place in the sequence is ignored, each event is followed by one of two other events each with a probability of 0.5. A random control condition was included where each event could be followed by any of the other 5, yielding transition probabilities of 0.2. When pairwise associations are probabilistically predictive, learning of transitional probabilities can speed reaction times (see also Jackson and Jackson, 1992; Stadler, 1992). Goschke found significant learning of the ambiguous sequences under distraction and argues that this result reflects learning of transition probabilities without the learning of context that would definitively specify the element at a particular place in the sequence. Nevertheless, he considers freedom from distraction to be necessary for building a representation that uses context to specify place in the sequence.

We have presented evidence that the nonattentional learning mechanism cannot learn sequences that are entirely composed of ambiguous associations. Nevertheless, significant learning of such sequences has been shown to occur under distraction (Keele and Jennings, 1992; Reed and Johnson, 1994). The results are inconclusive because distraction is unlikely to completely eliminate attention, and some nonattentional learning of ambiguous associations might also occur under greatly extended training. Furthermore, the difficulty of the secondary task can vary between experiments. Cohen et al (1990) showed that the difficulty of tone counting increased with the number of targets. Thus, it may be critical that Reed and Johnson found ambiguous sequence learning when 30% - 70% of the tones were targets, but Cohen et al. failed to find learning with 50% - 75% targets. Despite the contradictory nature of some of the results, there is consistent evidence that the ambiguous components of sequences containing both ambiguous and unique associations are learned under distraction (Curran and Keele, 1993; Frensch, Buchner, and Lin, 1994; Keele and Jennings, 1992). Thus, the hypothesized nonattentional mechanism must learn more than simple, pairwise associations. In our section on computational models we will consider candidate mechanisms.

Overall, the preceding experiments make a strong case for independent attentional and nonattentional learning systems. This again raises the question of whether these two systems differ in their learning capabilities. We have suggested that the attentional system has the capability of parsing sequences into chunks to build a hierarchic representation. Although such a conclusion is speculative, evidence to support it comes from a variety of explicit sequence-learning tasks cited in the introduction to this chapter (Restle and Burnside, 1972; Povel and Collard, 1982; and Gordon and Meyer, 1989), which showed that explicitly described sequences are coded in hierarchic form. Hierarchic coding in implicit memory has recently been investigated by examining the effects of exogenously parsing the sequence into sub-chunks with the insertion of temporal pauses at regular intervals. Stadler (1993) found that learning a completely ambiguous sequence was much better when the sequence was chunked in this way, even when subjects were not informed that a sequence was present.

There are at least two reasons why a hierarchic representation might depend on attention. First, hierarchic coding may require some kind of short-term memory process that preserves earlier portions of a sequence in order to chunk them with later portions. A distraction task might interfere with short-term memory, thereby preventing chunk formation (a similar idea has been expressed by Frensch, Buchner, and Lin, 1994; Frensch and Miner,

1994). A second possibility is that an attentional mechanism might serve a kind of place-keeping function. For a sequence with ambiguous associations, in order to know what event follows a current event, it may be necessary to keep track of the current position within the sequence, and attention may be necessary for that function. In another implicit learning paradigm, artificial grammar learning, distraction impaired the ability to learn positional information, lending some credence to such an idea (Dienes, Broadbent, and Berry, 1991).

Stadler (1995) has proposed an alternative explanation for the effects of distraction that does not assign an important role of "attention". Stadler suggests sequences are learned as unique runs (or chunks) of stimuli, and that the boundaries of these chunks are influenced by extraneous cues. For example, chunking patterns can be shaped by the insertion of temporal gaps at consistent places within the sequence so that insertion of random gaps disrupts learning (Stadler, 1993). According to this theory, consistent grouping allows the same chunks to be consistently encoded. Conversely, random grouping leads to encoding a large number of inconsistent chunks that are more poorly learned due to fewer repetitions. Because insertion of random gaps has effects that are very similar to the effects of tone counting, tone counting may merely interfere with the normal organization of the sequence (Stadler, 1995). Thus, Stadler suggests that transferring to and from conditions of distraction has deleterious effects because the organization of the sequence is changed. Further research is necessary to test the implications of these various theories of the effects of attention and distraction on sequence learning.

The empirical work on sequence learning reviewed thus far has provided a number of useful insights for our modular theory of sequence learning. First, although sequence learning clearly benefits from explicit knowledge (Cohen and Curran, 1993; Curran and Keele, 1993; Perruchet and Amorim, 1992; Willingham, et al., 1989), other work suggests that sequence learning does not require explicit knowledge (Nissen and Bullemer, 1987; Nissen, Knopman, and Schacter, 1987; Nissen, et al., 1989; Stadler, 1989, Reed and Johnson, 1994; Willingham, Greenley, and Bardona, 1993; Willingham, et al., 1989). Therefore, further work is necessary to distinguish between the mechanisms for implicit or unaware learning and those for explicit sequence learning. Work from our own lab suggests that, even when learning occurs implicitly, the learning that occurs under distraction is qualitatively different from learning that occurs when attention is fully available (Cohen, Ivry, and Keele, 1990; Curran and Keele, 1993). Finally, we must differentiate between the mechanisms for learning and representing sequences and mechanisms for activating the motor systems that are controlled by these representations (Keele et al., in press; Stadler, 1989). One way to investigate these questions is through the development of computational models.

COMPUTATIONAL EXPLORATIONS OF SEQUENCE LEARNING

Two important and related network models of sequence learning have been developed by Jordan (1986; 1995) and Elman (1990). We (Keele and Jennings, 1992) have explored Jordan's model to see whether it can account for our empirical results, and Cleeremans (1993) has similarly examined Elman's model. These particular models have provided two benefits. First, they have allowed us to formulate more precise ideas about possible meanings of terms like parsing, hierarchy, and chunking, and the role such factors play in sequence learning. Second, the models have suggested gaps in the existing data that must be filled before further computational progress can be made.

Jordan's model is a network of connections between input units, hidden units, and what we call prediction units (see Figure 8). The input units can be viewed as stimulus patterns that when processed through hidden units, produce output patterns on units that "predict" what the next response should be. We assume that the activation patterns on prediction units reflect the extent to which a particular response is primed.

Figure 8. Jordan's (1986; 1995) network model of sequential behavior. Not all connections are shown. Each plan and state unit connects to each hidden unit and each hidden unit connects to each prediction unit. Prediction errors, the difference between predicted and actual responses, are used to adjust weights by back propagation. The actual response is determined by a presented stimulus. The presented stimulus also feeds a single state unit with a weight of 1.0. Each state unit feeds back on itself with a fixed weight, μ.

The input layer consists of two segregated sets of units. One set, called plan units, retains an unchanging activation pattern over a sequence of stimuli. One might think of the plan units as representing a higher-order representation of a sequence to be performed, much like a concept of a word. The "word" representation as embodied in the plan units would remain activated until all the constituent phonemes are produced and a different "word" is activated. Thus, plans act as a high level node in a hierarchic arrangement.

The second set of input units are called state units. Activation of the state units is a function of the stimulus on the current trial, *t*, as well as the state units' activation on previous trials *(t-n)*. Stimulus positions are coded locally with a value of 1 passed to the current stimulus/state position, and 0 to all others. The state units also feed back upon themselves with a recurrent connection of weight $\mu < 1$. The result of these two inputs to each state unit is a representation of the stimulus that is influenced by the locations of prior stimuli. One state unit will be strongly activated by the current stimulus. That and other state units may have residual activation from past states. The rate of decay of past states depends upon the parameter setting of the recurrent connection with weight μ. If μ is low, less than about 0.2, the current stimulus tends to dominate, and there is little residual memory of events further back. If μ is high, greater than about 0.8, there is little loss of past states so that events of the distant past retain too great an influence. For intermediate values of μ, the state units provide a kind of moving window which represents the context in which stimuli occur. It is the maintenance of context that is critical for building associations between nonadjacent events.

The hidden units combine input from both plan and state units. The particular weightings assigned to the connections from state and plan units to the hidden units change as a function of learning. The hidden units provide input to the prediction units, and those weightings also change as a function of learning. The discrepancy between the actual response as determined by the stimulus (the "teacher") and the predicted response which is set to be the desired pattern of output is the source of error that propagates backward through the network resulting in gradual changes of weights until the network reliably produces the desired outputs.

The operation of the network occurs as follows: A pattern of activation appears on the plan units representing a sequence to be produced. Different sequences would have different plans. The state units, in the absence of a prior output, are all set at zero. The combination of plan input and zero input from the state units feeds through the network to produce the prediction of the first stimulus event. At that point in time, an external signal, such as a stimulus in a particular position, triggers the first response output in the series. Any prediction unit that does not match the response defines an error, and the error is then used by a back-propagation algorithm to adjust connection weights from state and plan units to hidden units and from hidden units to prediction units. The result of such weight adjustment is such that if exactly the same patterns of state and plan activity were to be fed through the network again, error would be reduced. At the next time step, the combination of plan units, which have remained fixed, and state units, one of which has been altered, is different than on the preceding time step. This new input pattern will lead to a different prediction than before as the information flows through the network.

The separate contributions of plan units and state units can be appreciated by considering a series of outputs that are identical to a particular place in a sequence. For example, in speech production of the word elegant vs. elephant, the first three phonemes are identical, and through that point, state units go through identical settings. What is it, therefore, that allows the system to correctly branch to the "g" in one word and the "ph" in the other? Outputs depend not on state units alone but on the conjunction of the plan and state units as mediated by connections to hidden units. Since the plan inputs are different for an intent to pronounce two different words, the conjunctions of plan and state are different, once the critical fourth phoneme is reached, and it is the difference in conjunction that results in different flow patterns.

Consider also a case in which identical stimuli occur at two different places in a sequence but are followed by different stimuli at those two places (e.g., 1213). How is the network able to accommodate that? Although two stimuli may themselves be identical, and therefore send identical input to the state units, the residual activity in the state units would typically differ, reflecting differing prior contexts for the two identical events. Thus, the total state unit pattern would differ for the two places in the sequence having identical stimuli, resulting in different predictions for what comes next.

The role of the plan units is somewhat different from the typical function of a node in a hierarchy. The plan units do not directly cause events to occur. Rather, plan units that remain constant over a series of events act concurrently with state units to jointly determine the next element in a series. Still, it is appropriate to think of the plan units as a level of representation higher in hierarchy than state units, because the plan units change less frequently and allow identical states to elicit plan-dependent predictions.

Jordan's model captures in computational form several of the concepts that we have invoked to explain our empirical results on sequence learning. Plan units can be viewed as implementing the concept of hierarchic representation or chunking. Parsing can be viewed as a process that resets state units to zero at the end or before the beginning of a sequence, thus marking its beginning and end. In computational form, one could imagine the last element of a sequence always to be a "null" element that causes resetting. One might suppose that

attentional distraction blocks parsing processes that discover starting and ending points and disables the plan systems that assign representations to chunks.

To determine whether these concepts are adequate to explain features of our data, we (Keele and Jennings, 1992) ran a series of simulations in which the model learned the same sequences that we presented to human subjects. We examined sequences of the type 132312..., in which all pairwise associations were ambiguous, as well as sequences with some uniquely occurring elements (e.g., events 2 and 4 in the sequence 132314...). The simulations involved blocks of 120 successive signals, i.e., 20 cycles through a sequence, and 10 blocks of trials were given. The first block of a trial could start at any position in the sequence. On each step in a series, the pattern of activation of prediction units was compared to the actual next stimulus and the discrepancies were used to modify connection weights in the network.

We ran simulations with three different versions of the Jordan model to assess the roles of parsing and plan-dependent representation. In one simulation, although plan units were present, they never changed. In that sense, the plan units represent nothing about a particular sequence. Moreover, although state units started at zero on the first trial of a block of 120 trials, after that they were free running so there was no demarcation of the end of a cycle through a sequence. These manipulations were intended to simulate nonattentional learning, where we supposed parsing and hierarchic representation are unavailable. In this mode, the system works as a relatively sophisticated associational system. We say associational, because the recurrent loops in the state units provide a decaying memory of past events, allowing such memories to participate in associational learning that spans intervening items.

The Jordan system stripped of representation and parsing was able to readily learn the sequences with uniquely occurring elements. Although the ambiguous sequences also were eventually learned, such learning was substantially slower. When parsing was added such that state units were reset to zero whenever a cycle of the sequence ended, there was little change in learning rate of the sequences that contained unique events, but learning of ambiguous sequences improved dramatically, becoming as rapid as the former. Similarly, if instead of parsing, plan-dependent representation was implemented, learning of ambiguous sequences improved to equal that of sequences with uniqueness. We implemented hierarchic representation simply by assigning one pattern of activation on the plan units to part of a sequence (e.g., for the 132 of the sequence 132312) and a different plan pattern for the other part of the sequence (i.e., 312).

Some insight into the mechanism by which the network learned the sequences was provided by an additional simulation. Again we stripped the system of parsing by not resetting state units at the end of a cycle, and eliminated hierarchic representation by not changing plan units for different parts of a sequence. In this case, we complicated an ambiguous sequence (e.g., 132312...) by adding a unique element, making the sequence longer and seemingly more difficult (e.g., 1323124.....). The strictly associational system learned the longer sequence with a unique element more readily than the shorter but completely ambiguous sequence. To see whether this outcome would also occur with *human* subjects under nonattentional conditions, we compared two groups of subjects. One group received longer sequences with one unique event and another group received shorter ambiguous sequences. Both groups were tested with the secondary distraction task of tone counting, which would presumably block parsing and hierarchic representations. The group receiving the longer sequence learned more readily than the group receiving the shorter sequence, confirming the prediction of the model. Indeed, in one replication of the experiment, the group receiving the shorter, ambiguous sequence did not learn at all.

Why is it that adding a unique element to an otherwise all-ambiguous sequence is beneficial? The explanation can be found in the nature of recurrent feedback to state units. Whenever a uniquely occurring stimulus appears, it activates a state unit on the next iteration. Although the activation of the state unit is partially renewed after each successive event, it gradually decays. The unique state event serves as a kind of marker that helps distinguish

events from one part of a sequence from otherwise identical events in another part of the sequence. Exactly the same function is supplied by altering the pattern on plan units at different parts of the sequence. That is, the different plans that accompany different parts of a sequence endow those parts with unique features that help disambiguate associations. Resetting of the state units, or what we call parsing, serves a similar function of disambiguating otherwise similar events.

A general lesson emerges from these simulations. One reason why hierarchic coding like that in the Jordan model is so beneficial is that it provides an auxiliary cue to disambiguate the same items in different sequential contexts. Consider once again the speech example of "elegant" versus "elephant". Despite the fact that in early portions, identical series of phonemes occur, co-occurrence of a plan embodied in plan units provides a disambiguation, allowing an associational machine to branch in appropriate directions.

An outstanding problem with our particular simulations is that we have not endowed the system with an ability to discover chunks and assign representations on its own. We have simply shown that if chunks are preassigned, or sequences are pre-parsed, then the general associational system has a much easier time learning the events within a chunk. Some other network simulations, in particular one by Cleeremans (1993a; Cleeremans and McClelland, 1991), do not have this limitation. Cleeremans' network has some similarities to a Jordan net, though it is based on a slightly different architecture, the serial recurrent network (SRN) developed by Elman (1990). Instead of having recurrent feedback of a state unit on itself, the SRN has recurrent feedback within a hidden layer. Hidden unit activation is determined by the current stimulus as well as the hidden unit activation on the previous trial. Thus, the hidden unit representation of a given stimulus is a graded function of the representation of previous stimuli. This system is able to learn sequences that have partial to complete ambiguity.

The explanation for why the Elman network can learn ambiguous sequences is similar to why the Jordan net learns, especially when the plan units are functioning. First of all, the hidden units capture some of the recent past history of a string of events. Such prior context helps disambiguate a sequence. Thus, in the above sequence, the event to follow position 3 can be disambiguated if a memory is retained of the item preceding 3. However, since the recurrent connections are themselves plastic, unlike Jordan's state units (see Cleeremans 1993a) this recurrent influence changes with learning. An analysis of the information being "learned" at the hidden layer reveals that some of the units come to represent not just the preceding item, but small clusters of preceding items. Thus, hidden units that represent chunks act much like plan units in the Jordan system to help disambiguate otherwise ambiguous pairwise associations. Importantly, the "chunks" in the hidden cells of the Elman system are self-discovered in the process of learning a sequence.

One assumption often made about secondary tasks is that they prevent the focusing of attention on a primary task and result in an increased signal to noise ratio in network connections. Cleeremans and McClelland therefore simulated the effects of distraction by adding noise to the hidden-unit input. Given that some of the hidden units eventually represent small subseries of events, adding noise to the hidden-unit input impairs the construction of chunks. They found that such added noise greatly impairs the learning of ambiguous sequences, but had a much reduced effect on sequences containing at least some unique associations. Such results were qualitatively similar to the empirical data of Cohen, Ivry and Keele (1990).

Despite the fact that the Cleeremans and McClelland model captures some aspect of the process by which chunks are discovered, it does not predict basic features of studies (Curran and Keele, 1993) involving transfer between attentional and nonattentional states. In his more recent work, Cleeremans (1993b) developed a "dual SRN" model that is able to simulate the empirical results. The model employs both a serial recurrent net, as described earlier, and a short-term memory buffer that has independent knowledge about a sequence when no distraction is present. This short-term buffer interacts with the basic network by

explicitly predicting each sequential element and allowing those predictions to modify the network's hidden-unit representation of the sequence. Put another way, the new model offered by Cleeremans still has two different knowledge systems, one for explicit knowledge and one for implicit knowledge. Rather than being strictly independent, however, the explicit knowledge can be an input source to the implicit system, though not vice versa.

Despite the architectural differences between Keele and Jennings' (1992) adaptation of the Jordan network and Cleereman's (1993b) dual network, both models provide similar insights relevant to our modular concept of sequence learning. Both models have a basic associative learning mechanism at the core. The representational capabilities of this associative mechanism can be enhanced when allowed to interact with higher-level processes. In the Jordan network these higher level processes include parsing via state-resetting at the end of a sequence, and hierarchic organization via plan units. In Cleereman's model, predictions generated by the short-term buffer constrain the hidden-unit representations of the associative system. Furthermore, the associative mechanisms of both models represent more than pairwise associations. Both allow for a kind of contextual representation of inputs such that the representation of an event is influenced by prior events. Such features allow ready learning of sequences that have a mixture of unique and ambiguous associations. In both models, however, the manner of learning higher level representations that allow acquisition of more complicated sequences – plans in the Jordan model and the explicit knowledge system in the Cleeremans model – is unspecified. All that can be said is that availability of such information facilitates sequence acquisition.

To this point there has been fruitful interaction between empirical discoveries about sequence learning and computational analysis. This interaction has indicated a need for additional empirical analysis that would test the underlying assumptions of different models. In particular, it appears that two future developments would be useful in guiding computational models. First, at the empirical level, we have insufficient evidence about the order in which different subparts of a sequence are learned. Secondly, the different computational models make different assumptions about the architecture of a sequence learning system. That is, each sequence learning system has a number of subparts, configured in different ways. Empirical analysis in sequence learning needs to be oriented toward further decomposition of the processes involved so that these processes can be incorporated into computational models.

POSSIBLE NEURAL SUBSTRATES FOR SEQUENCE LEARNING

Despite these needs for further work, the joint empirical and computational work has suggested brain systems or structures that might be involved in sequence learning. The motor-independence of sequential representation is consistent with the idea that there are distinct neural locales for sequential representation and conversion to motor activity. Similar suggestions have been advanced in studies of patients having ideomotor apraxia, a form of apraxia in which movement is intact and fluent, but inaccurate in the patterns produced (Heilman, Rothi, and Valenstein, 1982; Gonzales and Heilman, 1985). In making a salute, for example, movements may approximate salutes in some aspects but miss in others. Heilman and colleagues found such apraxic syndromes to occur following lesions either of posterior parietal cortex or of frontal cortex, the latter presumably involving areas of premotor cortex. However, if these same patients are asked to observe pairs of gestures one correctly performed and one poorly performed and indicate the correct gesture, patients with frontal damage perform well. Patients with parietal damage perform poorly. Thus, the patients with parietal lesions do poorly not only on motor production but on perceptual recognition; the patients with frontal lesions do poorly only on the production.

It seems plausible that the brain areas controlling sequential learning and performance may show a distribution of function that is similar to that seen in Heilman's apraxic patients. For visual-spatial actions, a parietal mechanism could subserve sequential learning and memory. In this respect it is useful to recall the patients with lesions in posterior parietal cortex described by Hillis and Caramazza (1988). These patients produced sequencing errors in either written spelling or oral spelling. The same mechanism that subserves sequential learning might also specify the location of future responses and interact with frontal mechanisms that specify articulators. Such fronto-parietal interactions have been hypothesized to subserve learning-dependent control of action, especially involving spatial tasks, in a number of domains (e.g., Fuster, 1993; Goldman-Rakic, 1990; Goodale, 1993; Passingham, 1993).

Our conclusion that sequential representation is effector independent leads us to speculate that at least visual-spatial sequences are represented in parietal cortex, whereas representations that control sequential performance are represented in frontal cortex. Frontal cortex may also contribute to sequence learning by functioning like plan units in Keele and Jennings' (1992) simulations. Rizzolatti and Gentilucci (1988) review single-cell work from their laboratory demonstrating that inferior regions of frontal cortex have properties suggestive of plans. Cells in inferior premotor cortex (inferior area 6) of the monkey become active when the monkey makes particular kinds of arm or mouth movements. Some cells become active during precision grasps involving finger and thumb; others become active for whole-hand grasps involving all fingers. Some cells are active if the monkey grasps with either the hand or the mouth. The cells involved in grasps do not become active if the hand is configured in similar ways but for purposes other than grasping. Many of the cells continue to fire throughout a series of actions that comprise a behavior. Thus, a "precision grasp cell" might start to fire as the monkey's arm is reaching toward a target, and continue to fire as the hand opens and then closes on the object. Such properties are reminiscent of the way plan units behave in the Jordan model, in which a plan represents a particular action sequence rather than specific components of the action and therefore remains constant throughout the execution of all of the components specified by the plan.

It is especially interesting in the work of Rizzolatti and Gentilucci that some neurons in inferior area 6 represent a similar grasp whether accomplished by the mouth or by the hand. Such results are in accord with our suggestions that a sequential representation is independent of the effector system of execution. Some cells in inferior area 6 are active not only during a monkey's grasp, but become active when the monkey observes a similar grasp performed by the experimenter (Rizzolatti, personal communication). Again, this suggests that the cells describe sequential events independently of the particular effector of execution.

Inferior area 6 of the monkey may be homologous to Broca's region in human cortex, long thought to be involved in human speech and language. Recently, Greenfield (1991) has suggested that Broca's area is specialized not for speech and language per se, but for hierarchic organization of event sequences. She points out, for example, that development of hierarchical control of various action sequences in infants occurs simultaneously with the development of hierarchical control of phonemes in speech. Moreover, hierarchical control of action and language reach the same relatively undeveloped stage in chimpanzees, compared to humans.

The observations by Rizzolatti and Gentilucci and arguments by Greenfield are consistent with a view that inferior portions of premotor cortex play a role in the chunking of events into sequences. In turn, one might speculate that such premotor regions interact with parietal regions to specify the particular events that make up a visuospatial sequence. In the context of Jordan's model, this would place plan units in premotor cortex, and state and possibly hidden units in parietal cortex. On the surface this scheme appears inconsistent with the earlier suggestion that apraxia due to parietal lesions results from a representational deficit while apraxia due to frontal lesions results from translation to particular motor effectors.

However, it is quite likely that different frontal regions or distinct distributed neural assemblies underlie these separate functions, especially given the variability of lesion sites and symptoms in patients with apraxia.

These speculations about the cortical loci of sequence representation differ from suggestions sometimes seen in the literature that sequence representation is largely subcortical. Previous research has shown that patients with basal ganglia dysfunction due to Parkinson's disease or Huntington's disease show impaired sequence learning (e.g., Ferraro, Balota and Conner, 1993; Jackson et al., in press; Knopman and Nissen, 1991; Willingham and Koroshetz, 1993). Our own suggestion is that the basal ganglia, rather than being a storage locus for sequence representation, are involved in sequence production. Although not a focus of this review, the hypothesis that the basal ganglia are involved in production has been explored in preliminary work in our laboratory (Hayes et al., 1995). When subjects were explicitly taught short sequences composed of two parts, patients with Parkinson's disease exhibited a substantial deficit at the transition point from one part to another whenever the identity of the second part differed from that of the first part. Such results suggest that the basal ganglia are part of a system that implements a shift from one sub-sequence representation to another, but that they might not be the site where sequence representation itself occurs. Such a hypothesis is in line with a broader hypothesis that the basal ganglia provide set-shifting functions across a range of domains.

Our long-term hope is that psychological analysis of the modules that make up sequential representation and production will form a reasonable basis for a neurological analysis. Our current thoughts are that the representation is distributed across posterior and frontal cortical regions and that additional frontal regions are involved in effector specification. The basal ganglia are part of an implementation system that allows progression through the representation in real time.

SUMMARY

Sequence learning may be comprised of a number of dissociable modules which subserve particular functions. Experimental evidence suggests that sequential representations are not tied to any particular effectors involved in executing responses, but exist at a more abstract level that specifies sequences of stimuli and/or responses rather than specifying specific actions.

The presence or absence of distraction, while not affecting independence of the representation from the effector system, does influence the kind of sequences that can be learned. Full attention allows the learning of more complex sequences that contain repeated events but in different orders in different portions of the sequence. Computational modeling suggests that attention enables mechanisms that parse a sequence to operate so that order within parts of the sequence can be represented. That is, attention allows hierarchic coding to take place.

In short, sequence learning and performance appears to be comprised of a complex of representational and control processes. Successful linking of brain mechanisms to sequence learning and production will require both an elaborated theory of cognitive processes and consideration of a diverse array of neural systems.

FOOTNOTE

Tim Curran is now on the faculty of Psychology at Case Western Reserve University in Cleveland, Ohio.

REFERENCES

Alexander, G. E., Crutcher, M. D., and DeLong, M. R., 1990, Basal ganglia thalamo-cortical circuits: Parallel substrates for motor control, oculomotor, "prefrontal" and "limbic" functions, *Prog. Brain Res.*, 85:119.

Berkenblit, M. B., and Feldman, A. G., 1988, Some problems of motor control, *J. Motor Behav.*, 20:369.

Berry, D.,1994, Implicit Learning: Twenty five years on. A tutorial, *in*: "Attention and Performance XV: Conscious and Nonconscious Information Processing", C. Umiltë and M. Moscovitch, eds., MIT Press, Cambridge, MA.

Cleeremans, A., 1993a, "Mechanisms of Implicit Learning: Connectionist Models of Sequence Processing", MIT Press, Cambridge, MA.

Cleeremans, A., 1993b, Attention and awareness in sequence learning, *in*: "Proceedings of the 15th Annual Conference of the Cognitive Science Society", Erlbaum , Hillsdale, NJ.

Cleeremans, A. and McClelland, J. L., 1991, Learning the structure of event sequences, *J. Exp. Psychol.: Gen.*, 120:235.

Cohen, A., and Curran, T., 1993, On tasks, knowledge, correlations, and dissociations: Comment on Perruchet and Amorim. *J. Exp. Psychol.: Learning, Memory, and Cognition*, 19:1431.

Cohen, A., Ivry, R. I., and Keele, S. W., 1990, Attention and structure in sequence learning, *J. Exp. Psychol.: Learning, Memory, and, Cognition*, 16:17.

Curran, T., and Keele, S. W., 1993, Attentional and nonattentional forms of sequence learning, *J. Exp. Psychol.: Learning, Memory, and Cognition*, 19:189.

Dienes, Z., Broadbent, D., and Berry, D., 1991, Implicit and explicit knowledge bases in artificial grammar learning, *J. Exp. Psychol: Learning, Memory, and Cognition*, 17:875.

Elman, J. L., 1990, Finding structure in time, *Cognit. Sci.*, 14:179.

Ferraro, F. R., Balota, D. A., and 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 type: A serial reaction time (SRT) investigation, *Brain and Cognition*, 21:163.

Frensch, P. A., Buchner, A., and Lin, J., 1994, Implicit learning of unique and ambiguous serial transactions in the presence and absence of a distractor task, *J. Exp. Psychol.: Learning, Memory, and Cognition*, 20:567.

Frensch, P. A., and Miner, C. S., 1994,Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning, *Memory and Cognition*, 5:95.

Fuster, J. M., 1993, Frontal lobes, *Curr. Opin. Neurobiol.*, 3:160.

Goldman-Rakic, P. S., 1990, Cellular and circuit basis of working memory in prefrontal cortex of nonhuman primates, *Prog. Brain Res.*, 5:325.

Gonzalez, L. J., and Heilman, K. M., 1985, Ideomotor apraxia: Gestural discrimination, comprehension and memory, *in*: "Neuropsychological Studies of Aapraxia", E. A. Roy, ed., North Holland Publishers, New York, NY.

Goodale, M. A., 1993, Visual pathways supporting perception and action in the primate cerebral cortex, *Curr. Opin. Neurobiol.*, 3:578.

Gordon, P. C., and Meyer, D. E. , 1987, Control of serial order in rapidly spoken syllable sequences, *J. Memory Language*, 26:300.

Grafton, S. T., Hazeltine, E., and Ivry, R., 1995, Functional mapping of sequence learning in normal humans, *J. Cognit. Neurosci.*, in press.

Greenfield, P. M., 1991, Language, tools and brain: The ontogeny and phylogeny of hierarchically organized behavior, *Behav. Brain Sci.*, 14:531.

Heilman, K. M., Rothi, L. J., and Valenstein, E., 1982, Two forms of ideomotor apraxia, *Neurology*, 32:342.

Hillis, A. E., and Caramazza, A., 1988, The graphemic buffer and attentional mechanisms, Report no. 30, Cognitive Neuropsychology Laboratory, Johns Hopkins University, Baltimore, MD.

Ivry, R. I. and Gopal, H. S., 1992, Speech production and perception in patients with cerebellar lesions, *in*: "Attention and Performance XIII", D. Meyer and S. Kornblum, eds., MIT Press, Cambridge, MA.

Ivry, R. I. and Keele, S. W., 1989, Timing functions of the cerebellum, *Cognit. Neurosci.*, 1:134.

Jackson, G., and Jackson, S., 1992, Sequence structure and sequential learning: The evidence from aging reconsidered, Technical Report No. 92-9, Institute of Cognitive and Decision Sciences, University of Oregon, Eugene, OR.

Jackson, S. R., Jackson, G. M., Harrison, J., Henderson, L., and Kennard, C., 1995, Serial reaction time learning and Parkinson's disease: Evidence for a procedural learning deficit, *Neuropsychologia*, 33:577.

Jordan, M. I., 1986, Serial order: A parallel distributed processing approach, ICS Report 8604, Institute for Cognitive Science, University of California, San Diego, La Jolla, CA.

Jordan, M. I., 1995, The organization of action sequences: Evidence from a relearning task. *J. Motor Behav.* 27, 179-211.

Keele, S. W., Cohen, A., and Ivry, R. I., 1990, Motor programs: Concepts and issues, *in*: "Attention and Performance XIII", M. Jeannerod, ed., Lawrence Erlbaum Associates, Hillsdale, NJ.

Keele, S. W., and Ivry, R., 1990, Does the cerebellum provide a common computation for diverse tasks: A timing hypothesis, *in*: "The Development and Neural Bases of Higher Cognitive Function", A. Diamond, ed., *Ann. NY Acad. Sci.*, 608:179.

Keele, S. W., and Jennings, P., 1992, Attention in the representation of sequence: Experiment and theory, *Human Movement Science*, 11:125.

Keele, S. W., Jennings, P., Jones, S., Caulton, D., and Cohen, A., 1995, On the modularity of sequence representations, *J. Motor Behav.* 27:17.

Knopman, D., and Nissen, M. J., 1991, Procedural learning is impaired in Huntington's disease: Evidence from the serial reaction time task, *Neuropsychologia*, 29:245.

Lavond, D. G., Lincoln, J. S., McCormick, D. A., and Thompson, R. F., 1984, Effects of bilateral lesions of the dentate and interpositus nuclei on conditioning of heart-rate and nictitating membrane/eyelid responses in the rabbit., *Brain Res.*, 305:323.

MacKay, D. G., 1982, The problem of flexibility and fluency in skilled behavior, *Psychol. Rev.*, 89:483.

MacKay, D. G., 1987, "The Organization of Perception and Action", Springer-Verlag, New York, NY.

Mayr, U., 1994, "Spatial attention and implicit sequence learning: Evidence for independent learning of spatial and nonspatial sequences, Technical Report. 94-13, Institute of Cognitive and Decision Sciences, University of Oregon, Eugene, OR.

Mesulam, M.-M., 1985, Patterns in behavioral neuroanatomy: Association areas, the limbic system, and hemispheric specialization, *in*: "Principles of Behavioral Neurology", M.-M. Mesulam, ed., F. A. Davis Company, Philadelphia, PA.

Mesulam, M.-M., 1990, Large-scale neurocognitive networks and distributed processing for attention, language, and memory, *Ann. Neurol.*, 28:597.

Nissen, M. J., and Bullemer, P., 1987, Attentional requirements of learning: Evidence from performance measures, *Cognit. Psychol.*, 19:1.

Nissen, M. J., Knopman, D. S., and Schacter, D. L., 1987, Neurochemical dissociation of memory systems, *Neurology*, 37:789.

Nissen, M. J., Willingham, D., and Hartman, M., 1989, Explicit and implicit remembering: When is learning preserved in amnesia, *Neuropsychologia*, 27:341.

Pascual-Leone, A., Grafman, J., Clark, K., Stewart, M., Massaquoi, S., Lou, J.-S., and Hallett, M., 1993, Procedural learning in Parkinson's disease and cerebellar degeneration, *Ann. Neurol.*, 34:594.

Passingham, R. E., 1993, "The Frontal Lobes and Voluntary Action", Oxford University Press, Oxford, UK

Perrett, S. P., Ruiz, B. P., and Mauk, M. D., 1993, Cerebellar cortex lesions disrupt learning- dependent timing of conditioned eyelid responses, *J. Neurosci.*, 13:1708.

Perruchet, P., and Amorim, M., 1992, Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation, *J. Exp. Psychol.: Learning, Memory, and Cognition*, 18:785.

Povel, D. J., and Collard, R., 1982, Structural factors in patterned finger tapping, *Acta Psychol.*, 52:107.

Raibert, M. H., 1977, Motor control and learning by the state space model, Technical Report AI-M-351, NTIS AD-A026-960, Massachusetts Institute of Technology, Cambridge, MA.

Reber, A. S., 1989, Implicit learning and tacit knowledge, *J. Exp. Psychol.: Gen.*, 118:219.

Restle, F., and Burnside, B. L., 1972, Tracking of serial patterns, *J. Exp. Psychol.*, 95:299.

Rizzolatti, G., Gentilucci, M., 1988, Motor and visual-motor functions of the premotor cortex, *in*: "Neurobiology of Neocortex", P. Rakic and W. Singer, eds., Wiley, New York, NY.

Rosenbaum, D. A., 1987, Successive approximations to a model of human motor programming, *in*: "The Psychology of Learning and Motivation", G. Bower, ed., Academic Press, New York, NY.

Rozin, P., 1976, The evolution of intelligence and access to the cognitive unconscious, *in*: "Progress in Psychobiology and Physiological Psychology", J. M. Sprague and A. N. Epstein, eds., Academic Press, New York, NY.

Shanks, D. R., and St. John, M. F., 1994, Characteristics of dissociable human learning systems, *Behav. Brain Sci.*, 17:367.

Stadler, M. A., 1989, On the learning of complex procedural knowledge, *J. Exp. Psychol.: Learning, Memory, and Cognition*, 15:1061.

Stadler, M. A., 1992, Statistical structure and implicit serial learning, *J. Exp. Psychol.: Learning, Memory, and Cognition*, 18:318.

Stadler, M. A., 1993, Implicit serial learning: Questions inspired by Hebb (1961), *Memory and Cognition*, 21:819.

Stadler, M. A., 1995, The role of attention in implicit learning, *J. Exp. Psychol.: Learning, Memory, and Cognition*, 21:674.

Thompson, R. F., 1986, The neurobiology of learning and memory, *Science*, 233:941.

Thompson, R. F., 1990, Neural mechanisms of classical conditioning in mammals, *Philos. Trans. R. Soc. Lond.*, B329:161.

Willingham, D. B., 1992, Systems of motor skill, *in*: "Neuropsychology of Memory", L. R. Squire and N. Butters, eds., The Guilford Press, New York, NY.

Willingham, D. B., Greenley, D. B., and Bardona, A. M., 1993, Dissociation in a serial response time task using a recognition measure: Comment on Perruchet and Amorim (1992), *J. Exp. Psychol.: Learning, Memory, and Cognition*, 19:1424.

Willingham, D. B., and Koroshetz, W. J., 1993, Evidence for dissociable motor skills in Huntington's disease patients, *Psychobiol.*, 21:173.

Willingham, D. B., Nissen, M. J., and Bullemer, P., 1989, On the development of procedural knowledge, *J. Exp.Psychol.: Learning, Memory, and Cognition*, 15:1047.

Wright, C. E., 1990, Generalized motor programs: Reexamining claims of effector independence in writing, *in*: "Attention and Performance XIII", M. Jeannerod, ed., Lawrence Erlbaum Associates, Hillsdale, NJ.

Wright, C. E. and Lindemann, P., 1993, Effector independence in hierarchically structured motor programs for handwriting. Presented at 34th Annual Meeting of the Psychonomic Society, Washington, DC.

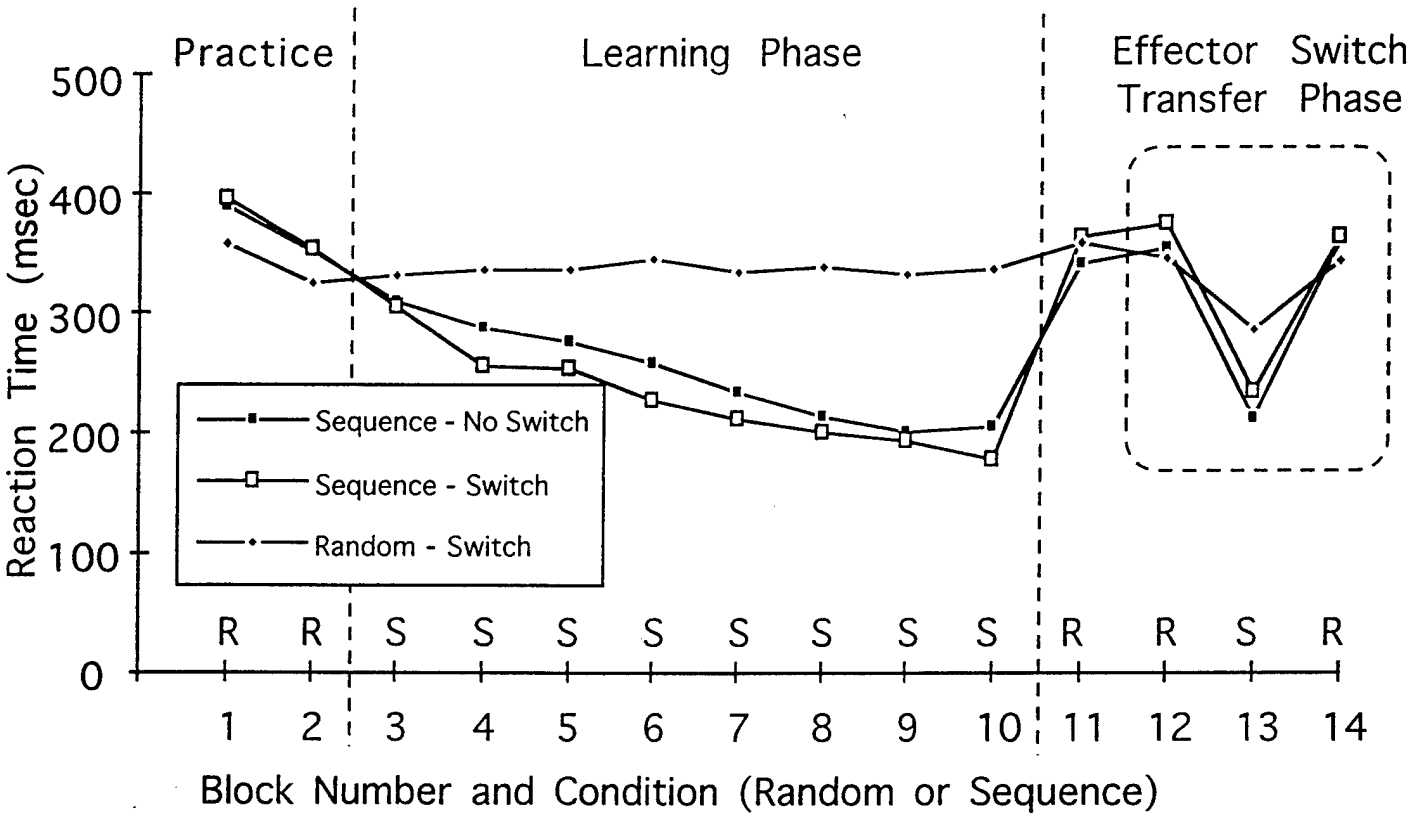
Practice
Learning Phase
Effector Switch
Transfer Phase
500
400
300
200
100
0
Reaction Time (msec)
Sequence - No Switch
Sequence - Switch
Random - Switch
R R S S S S S S S S R R S R
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Block Number and Condition (Random or Sequence)

Figure 1

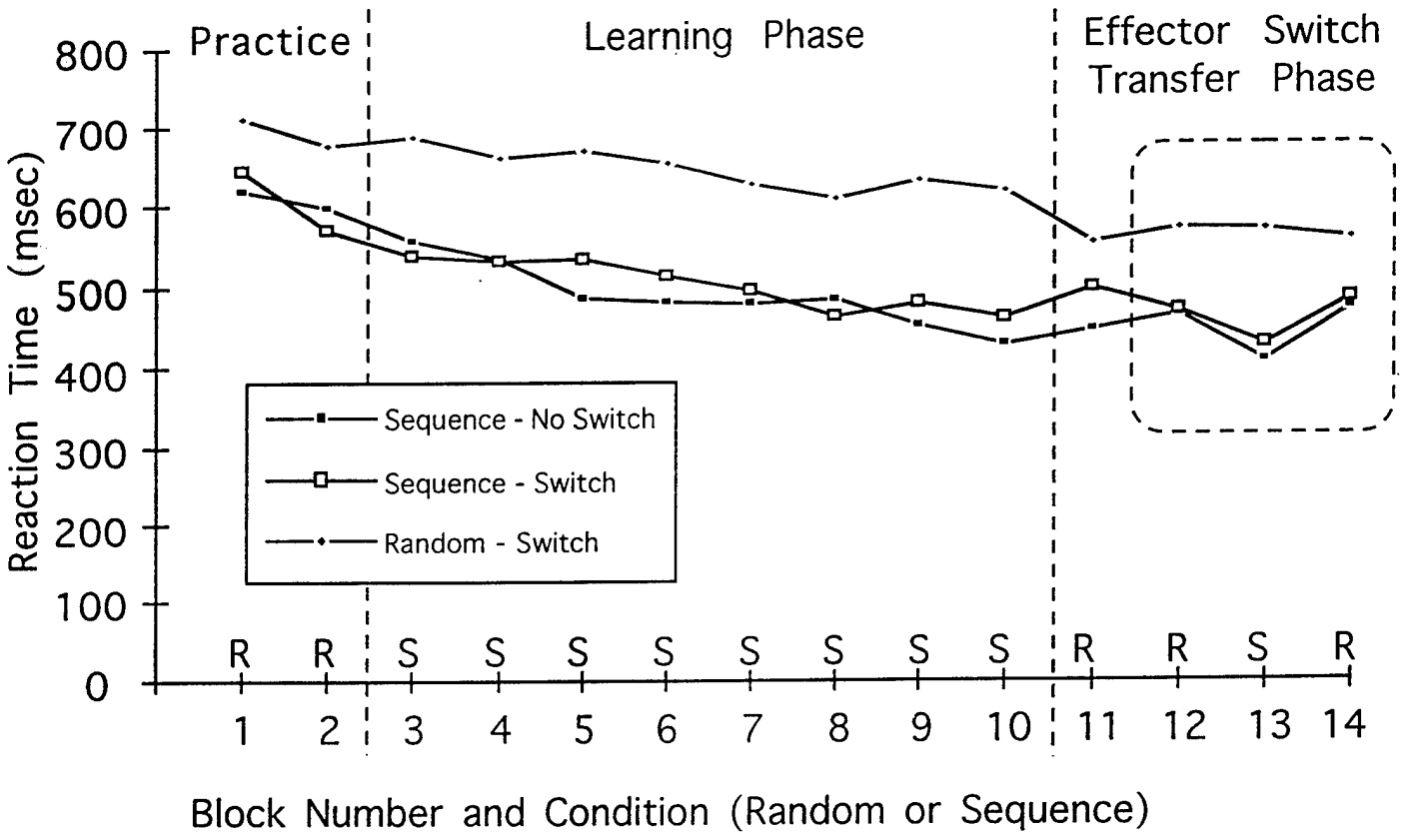
Practice
Learning Phase
Effector Switch Transfer Phase
Reaction Time (msec)
800
700
600
500
400
300
200
100
0
Sequence - No Switch
Sequence - Switch
Random - Switch
R R S S S S S S S S R R S R
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Block Number and Condition (Random or Sequence)

Figure 2

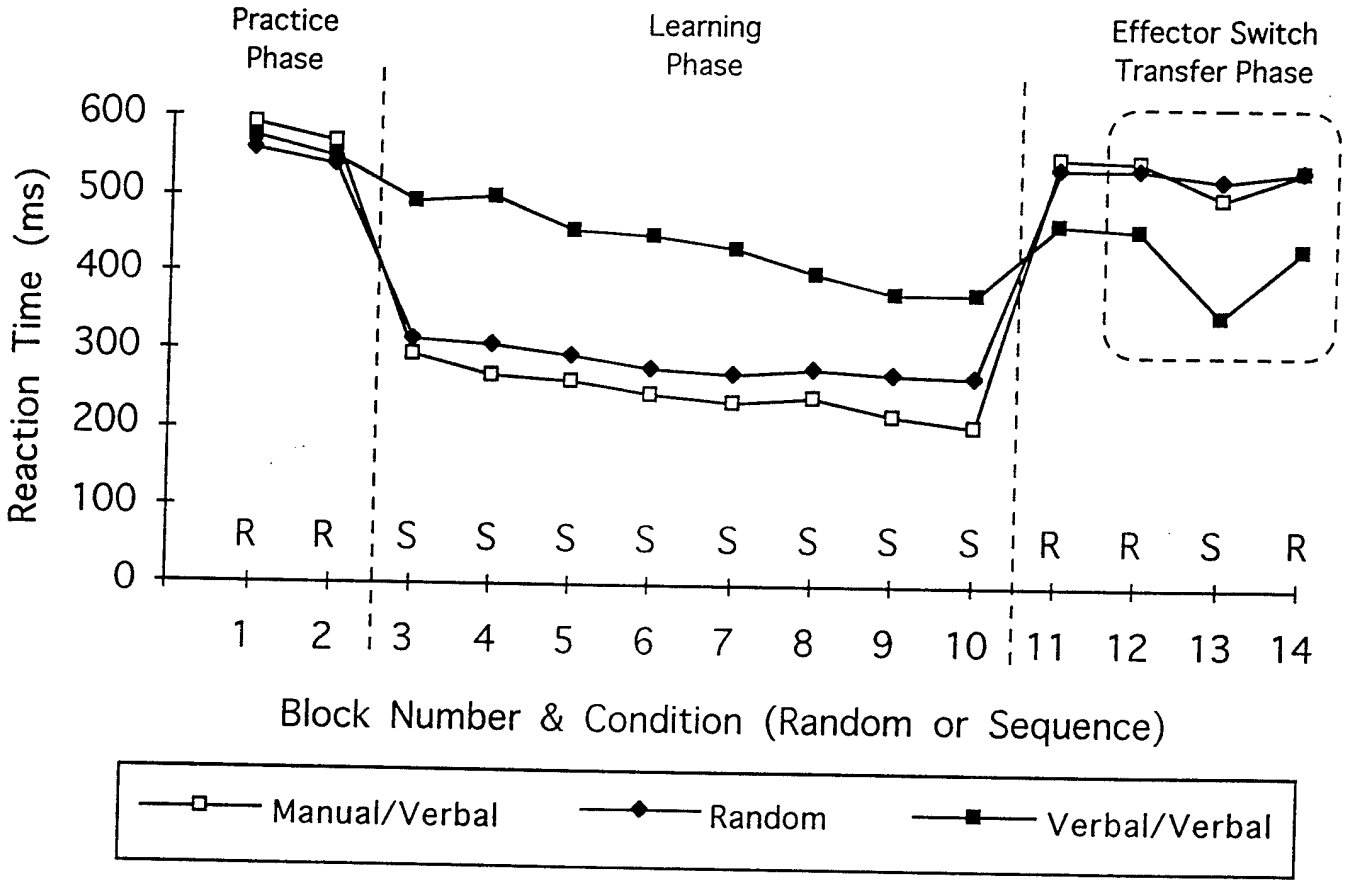
Practice Phase
Learning Phase
Effector Switch Transfer Phase
Reaction Time (ms)
600
500
400
300
200
100
0
R R S S S S S S S S R R S R
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Block Number & Condition (Random or Sequence)
Manual/Verbal
Random
Verbal/Verbal

Figure 3

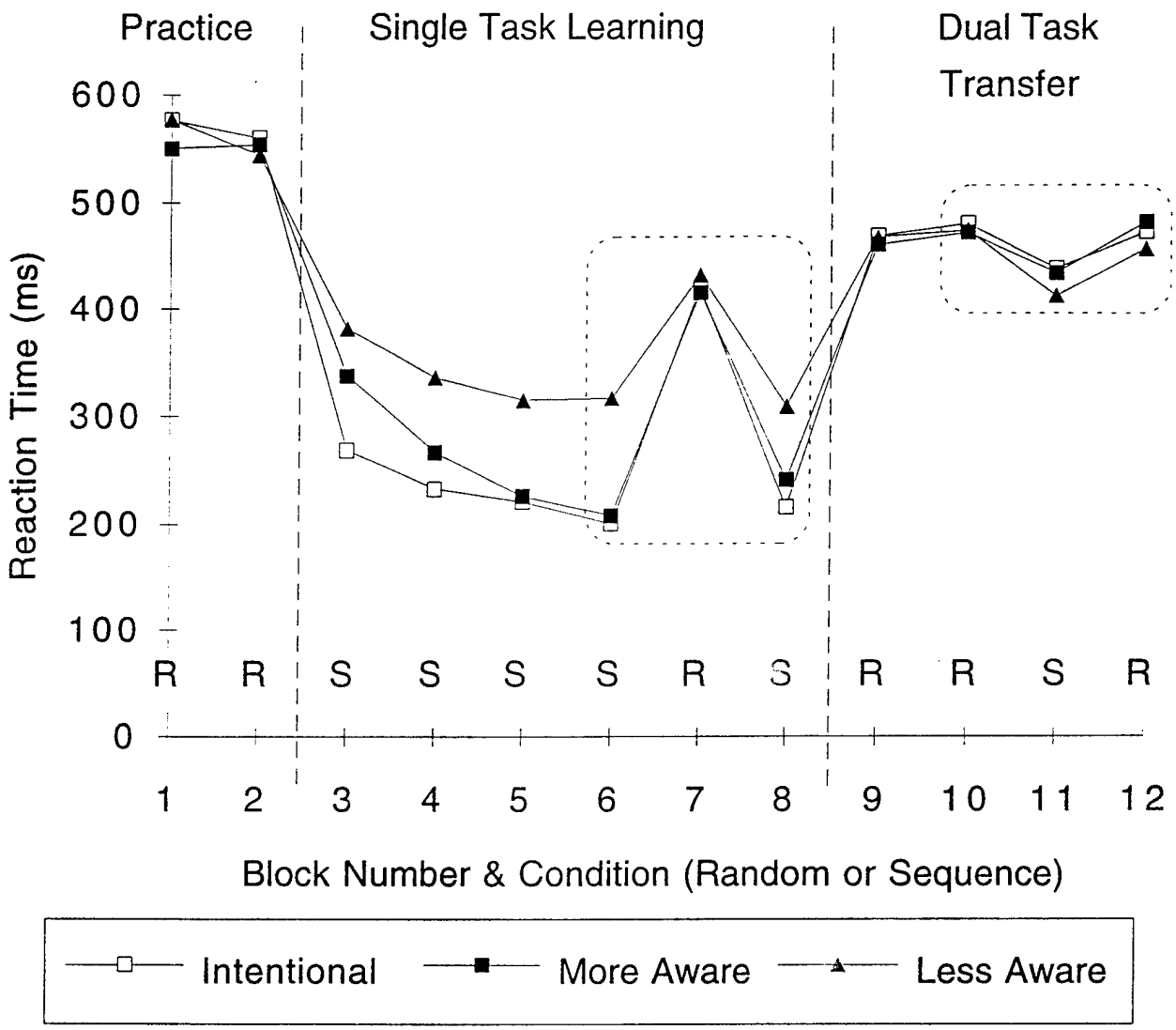
Practice
Single Task Learning
Dual Task
Transfer
600
500
400
300
200
100
0
Reaction Time (ms)
R R S S S S R S R R S R
1 2 3 4 5 6 7 8 9 10 11 12
Block Number & Condition (Random or Sequence)
Intentional
More Aware
Less Aware

Figure 4

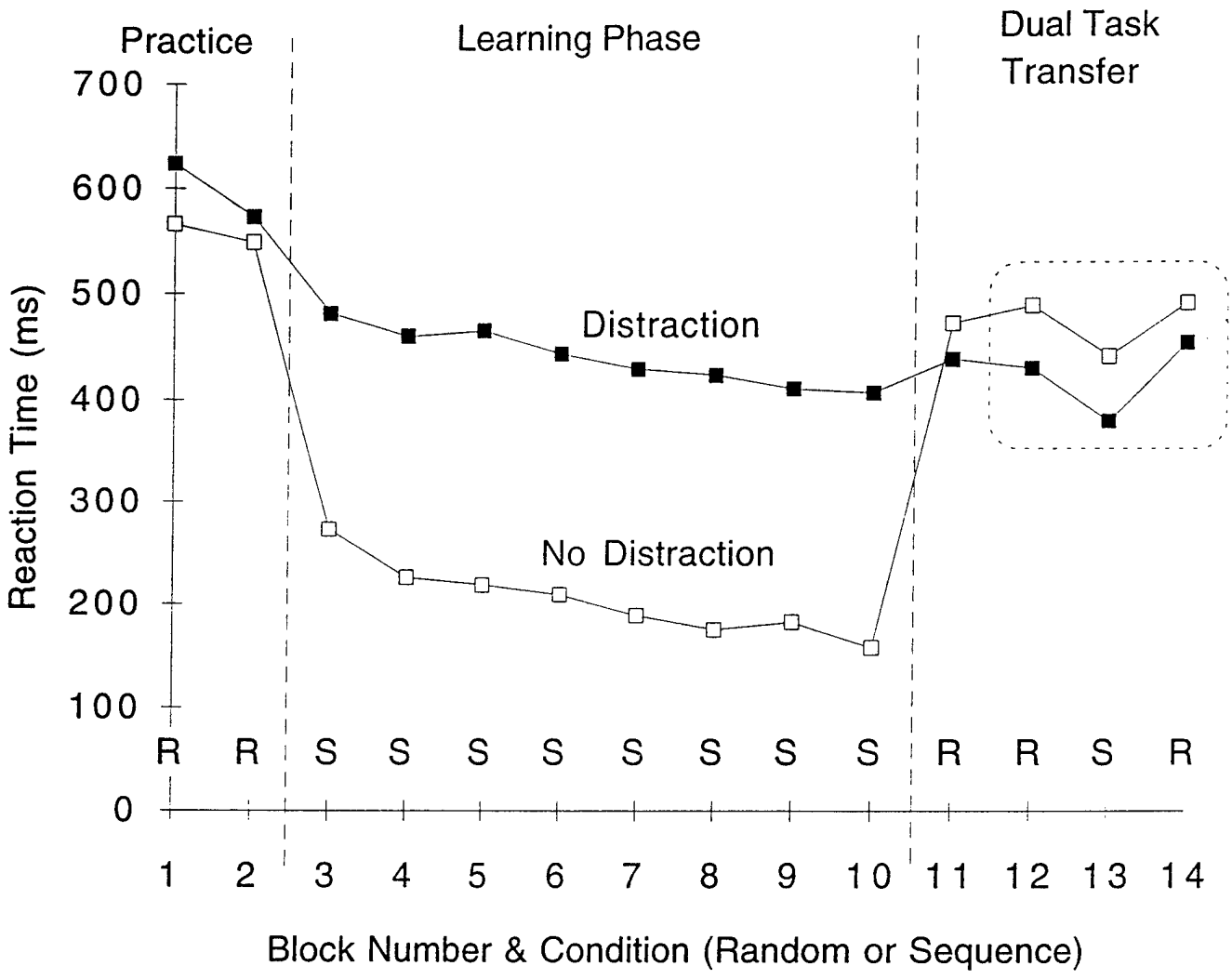
Practice
Learning Phase
Dual Task
Transfer
700
600
500
400
300
200
100
0
Reaction Time (ms)
Distraction
No Distraction
R R S S S S S S S S R R S R
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Block Number & Condition (Random or Sequence)

Figure 5

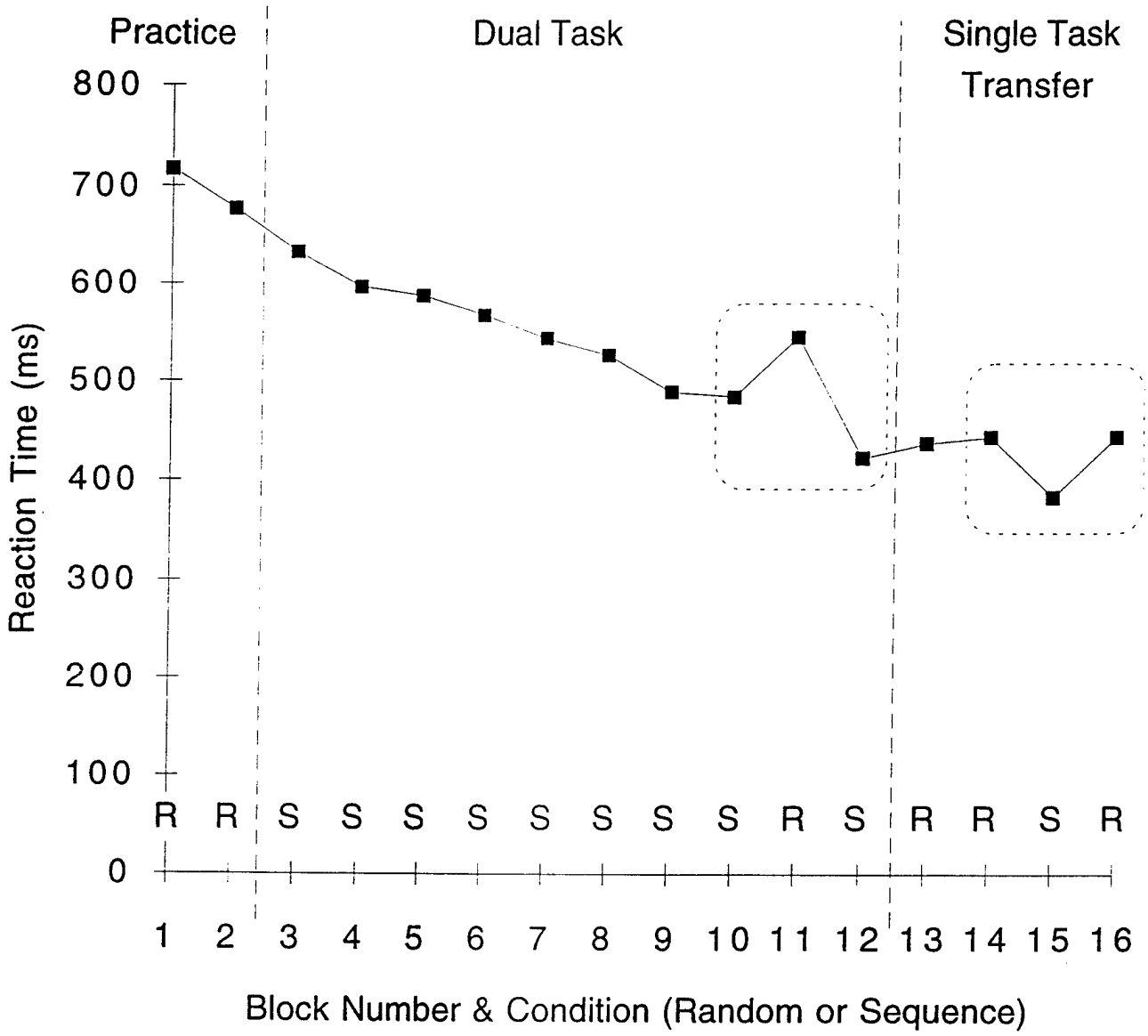
Practice
Dual Task
Single Task
Transfer
800
700
600
500
400
300
200
100
0
Reaction Time (ms)
R R S S S S S S S S R S R R S R
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Block Number & Condition (Random or Sequence)

Figure 6

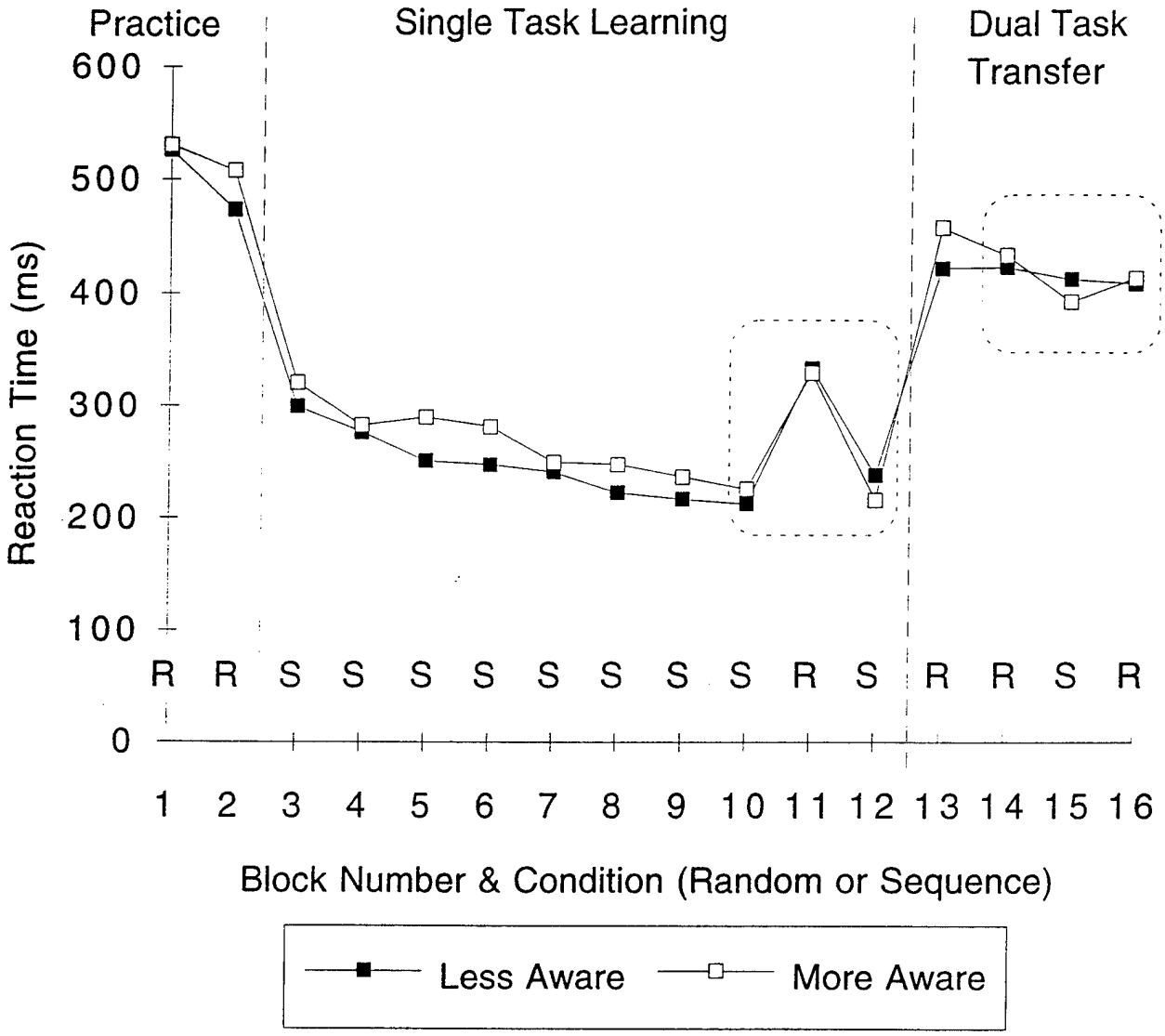
Practice
Single Task Learning
Dual Task
Transfer
600
500
400
300
200
100
0
Reaction Time (ms)
R R S S S S S S S S R S R R S R
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Block Number & Condition (Random or Sequence)
Less Aware
More Aware

Figure 7

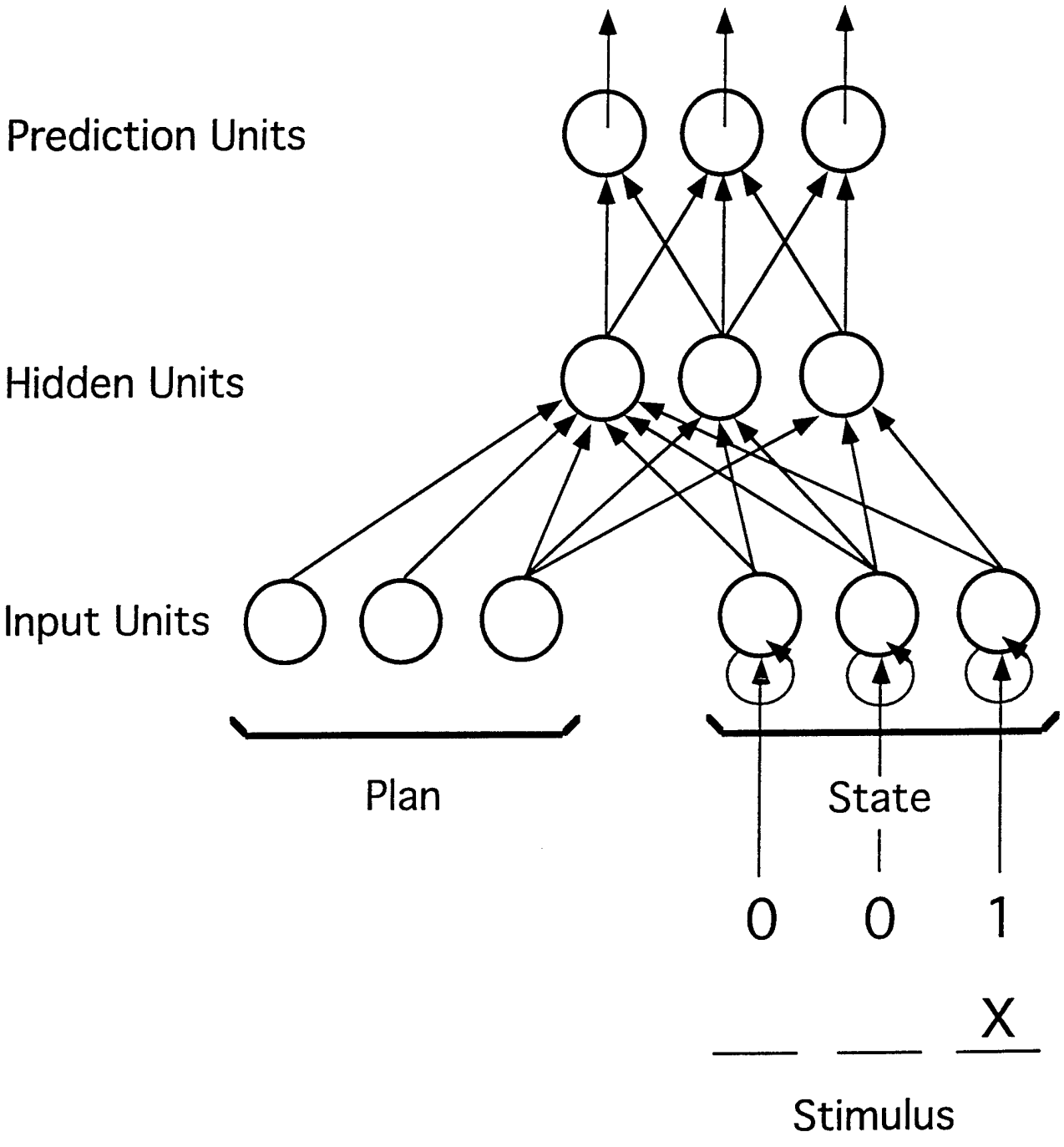
Prediction Units
Hidden Units
Input Units
Plan
State
0
0
1
X
Stimulus

Figure 8