

Parallel Models of Serial Behaviour: Lashley Revisited

George Houghton & Tom Hartley

Dept. of Psychology
University College London
London, WC1 6BT
U.K.

g.houghton@psychol.ucl.ac.uk

Copyright (c) George Houghton & Tom Hartley 1996

PSYCHE, 2(25), February 1995

<http://psyche.cs.monash.edu.au/v2/psyche-2-25-houghton.html>

KEYWORDS: serial order, response competition, neural networks, competitive queuing, response chaining, spelling, verbal short-term memory, syntax.

ABSTRACT: In 1951, Lashley highlighted the importance of serial order for the brain and behavioural sciences. He considered the response chaining account untenable and proposed an alternative employing parallel response activation and "schemata for action". Subsequently, much has been learned about sequential behaviour, particularly in the linguistic domain. We argue that these developments support Lashley's picture, and recent computational models compatible with it are described. The models are developed in a series of steps, beginning with the basic problem of parallel response competition and its possible resolution into serial action. At each stage, important limitations of the previous models are identified and simple additions proposed to overcome them, including the provision of learning mechanisms. Each type of model is compared with relevant data, and the importance of error data is emphasized. Taken together, the models constitute a unified approach to serial order which has achieved considerable explanatory success across disparate domains.

1. Introduction: The Problem of Serial Order

1.1 In a well-known article published in 1951, "The Problem of Serial Order in Behavior", Karl Lashley proposed that the problem of how behavioural sequences are produced should be of central concern to the neuropsychologist and physiologist. He pointed out that sequential organization is central to much of animal and human behaviour, from locomotion, through reaching and grasping to language and the control of logical reasoning. This organization could not be attributed to moment by moment responding to a serially ordered environment, but rather depends upon internal organizing principles by which the animal controls its own behaviour. Surveying the then available

ideas, Lashley concluded that neither the neurosciences nor psychology had much insight to offer into the problem. The "only strictly physiological theory" to have been explicitly formulated was associative chaining theory, in which it is postulated that each element in a series of actions provides the excitation of the next (various examples of such theories, which remain current, are discussed below). From consideration of a variety of qualitative data, particularly regarding speech errors, Lashley came to the conclusion that such theories are untenable. He postulated instead that the production of serial behaviour involves the parallel activation of a set of actions, which together comprise some "chunk", so that responses are internally activated before being externally generated. This activation, in itself, does not contain the serial ordering of the actions. Superimposed on this activation is some kind of independent ordering system, a "schema for action", which selects which response, of those activated, to produce at which time. Unfortunately, Lashley was able to progress no further, writing that,

[I]ndications ... that elements of the [sequence] are ... partially activated before the order is imposed upon them in expression suggest that some scanning mechanism must be at play in regulating their temporal sequence. The real problem, however, is the nature of the selective mechanism by which the particular acts are picked out in this scanning process and to this problem I have no answer.

1.2 In the years following Lashley's article, the "cognitive revolution" got underway, born of the union of theoretical linguistics (Chomsky, 1957) and the computer metaphor for the mind developed in artificial intelligence. In these subjects, the problem of serial order is solved technologically. Theoretical linguistics, which only concerns itself with the internal representation of serial order ("competence") and not with its execution ("performance"), avails itself of such formal objects as ordered sets, strings etc. as primitives, from which to build descriptions of grammars and other abstract objects. In artificial intelligence and computer science, analogous objects plus recursive serial processing are provided by computer programming languages. In such a context, serial order per se will not appear to be any kind of problem at all. Thus, although these devices have never been defended or tested on empirical grounds, their availability and computational power suffice to obscure the fact that cognitive science has no (neuro-) psychologically grounded theory of serial order. In neurophysiology and psychology, the problem has been largely ignored or workers have fallen back on the very position that Lashley attacked, associative chaining theory. More recently however, there are signs that the problem of serial order may once again be being taken seriously in its own right. In cognitive science, the rise in the importance of neuropsychological and physiological data, coupled with the widespread use of neural network models, has led to the rethinking of many basic issues, and to the questioning of the uncritical use of classical symbol processing architectures. Any abandonment of the symbolic primitives mentioned above immediately throws the serial order problem to the fore (Houghton, 1990; Shallice, Glasspool & Houghton, in press). Additional impetus for the development of more biologically plausible models of serial order is provided by the increased interest in the problem in the neurosciences (Aldridge, Berridge, Herman, & Zimmer, 1993; Berridge & Wishaw, 1992; Colombo, Eickhoff & Gross, 1993; Kermadi, Jurquet, Arzi & Joseph, 1993; Paulesu, Frith & Frackowiak, 1993), and in animal learning studies (Fountain, Henne & Hulse, 1984; Terrace, 1991).

1.3 The ubiquity of serial order in behaviour, from something as simple as an eyeblink to the performance of a piano sonata, leads us to raise the issue of what range of behaviours a given model of serial behaviour should be expected to apply to, as it is quite possible that numerous different solutions to the serial order problem have evolved. All animals exhibit some degree of endogenous temporal structure in their behaviour (e.g., the different gait patterns in quadrupedal, walking, trotting and galloping). In some cases this sequential organisation can be quite complex e.g., in grooming sequences (Berridge & Wishaw, 1992), or bird song (Konishi, 1985), and exhibit a significant degree of learning (Marler, 1991). Is it reasonable to search for a global set of principles applicable to all such cases, or are local, task-specific solutions more likely? We believe that at present levels of understanding, it would be premature to attempt to answer this question decisively. We suggest that theoretical investigations should therefore be concentrated on specific, well-studied, classes of behaviours in particular animals (though potential insights from related studies should not be ignored). Empirically successful theories developed in this way can then be compared for the presence or absence of common principles. Accordingly, the models described in the current paper are all concerned with the voluntary production of sequences of actions by human subjects, where the same actions may occur in many different orders. Although at various points we discuss action sequencing in general, the models are largely motivated by data from studies of linguistic behaviour in speaking, spelling and typing. Even within this limited domain, the emphasis is on analyses of recall error data (from both normal and impaired subjects, and from short- and long-term memory), including the distribution of errors as a function of such variables as serial position and the familiarity of the target sequence. Error data from such sources were central to Lashley's (1951) original argument for the need for a new model of serial behaviour, and have continued to play a central role in constraining theories (MacNeilage, 1964; Mackay, 1970, 1972; Dell, 1986; Henson, Norris, Page & Baddeley, in press).

1.4 In the current terminology of the memory literature, such data would appear to mainly involve "explicit memory" tasks, in which remembered information is consciously used to guide behaviour. Subjects in serial recall tasks certainly know they are intended to reproduce the target list, and effortful concentration of attention is required to correctly repeat lists of more than a few items long. Speaking and writing in a particular language require correct, explicit, retrieval of the phonological and orthographic forms of words, and people know whether or not they possess such knowledge (though we certainly would not claim that they know what form this knowledge has or how it is they use it to guide their behaviour). Recently there has been increasing interest in the possibility of "implicit" sequence learning, whereby subjects show evidence in serial reaction time (SRT) tasks of learning properties of repeated sequences that they cannot explicitly recall or state (Nissen & Bullemer, 1987. See Curran, this volume, for a neurophysiologically-oriented review of this literature). What kind of knowledge this form of learning results in (or how it is put to use) is not clear from available SRT data (Jackson, Jackson, Harrison, Henderson & Kennard, in press). We therefore restrict our attention to models of explicit serial recall, by which we mean those cases in which the subject voluntarily uses what has been learned to attempt to produce a target sequence of actions.

1.5 Below, we review a number of relatively recent computational models of serial order which are compatible with Lashley's insights (and which we refer to generally as "competitive queueing" models). Three related classes of model are built up in a series of steps, beginning with the basic problem of parallel response competition and its possible resolution into serial action. At each stage, important limitations of the previous models are identified and simple additions proposed to overcome them, including the provision of learning mechanisms. Each type of model is compared with relevant data, and the importance of constraints from error data is emphasized. The final type of model discussed incorporates "schemata for action".

1.6 By way of contrast, we first consider the characteristics of the associative chaining approach to serial order, which Lashley considered "doomed to failure". Rosenbaum (1991; p. 80) states that "It is difficult to introduce response-chaining theory without appearing to treat it as a straw man." Our reason for discussing it here is that, in one guise or another, it seems to be always with us (e.g., Lewandowsky & Murdoch, 1989). Indeed, recent neural network models of serial order appear to have breathed new life into the "straw man", as they generally depend, at least in part, on the formation of associations between successive actions (Amit, Sagi, & Usher, 1990; Ans, Coiton, Gilhodes, & Velay, 1994; Bairaktaris, 1992; Dehaene, Changeux, & Nadal, 1987; Jordan, 1986).

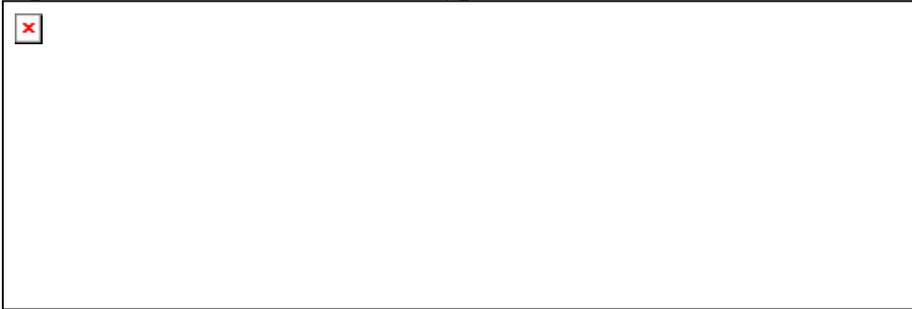
2. Serial Order and Associative Chaining

2.1 Some of the earliest psychological accounts of serial order postulated that action sequences were represented as chains made up of unidirectional S-R links. The appeal of this type of account is its simplicity; it requires nothing more than a representation of the items themselves and the links between them. Retrieval of a sequence is achieved by tracing a path through the links. One of the central difficulties Lashley (1951) identified with such a model is the handling of sequences containing repeated items. In such sequences, a stimulus action is associated with more than one response, and chaining models provide no mechanism for choosing between different associative links. This is illustrated in Figure 1a, which shows the associative links necessary to represent the sequence "E, V, E, R, Y". "E" must be linked to both "V" and "R", so that it is not clear which item follows the first instance of "E". This problem arises again when the second instance of "E" is realized, and there is the potential for endless looping through the first part of the sequence, without ever reaching the final item. The problem is exacerbated when one considers a single associative structure containing representations of a large number of sequences, for example sequences of speech sounds making up the familiar words in a mental lexicon. If the same elements are to be used to stand for the /t/, /a/, and /k/ in tack, cat and act, then the chains will be linked together to form a network in which the underlying serial structure of any single word is obscured by links between them.

Figure 1

Simple forms of Associative Chaining (AC): (a) AC has problems with sequences containing repeats. the attempt to represent every leads to uncertainty as to which link from e should be followed; (b) Wickelgren's Context-Specific Coding overcomes

these difficulties, but only at the expense of using entirely different tokens to represent instances of the same type.



2.2 To overcome this difficulty, Wickelgren (1969) suggested a form of context-sensitive chaining. Elementary actions were represented by different tokens depending on their immediate context. For example the sequence "E, V, E, R, Y" would be represented as the set of tokens \$Ev , eVe, vEr, eRy, rY\$, where \$ is an end marker, and lower case letters represent local context. Representations based on Wickelgren's idea have proved popular with some neural network modellers (Rumelhart & McClelland, 1986; Seidenberg & McClelland, 1989). Because the approach uses tokens (e.g., standing for particular instances of each "E" in the example) rather than types (standing for the category of actions designated "E"), a chain involving a repeated action can be represented without linking one stimulus to more than one response. In Figure 1b. it is clear that by following the stimulus-response chain from left to right, the target sequence "E, V, E, R, Y" can be generated. However, the token based form of representation is immediately unappealing, because it fails to capture any relationship between different instances of the same item in a sequence. In the above example, the two "E"s are as different from one another as they are from the other letters in the sequence. In fact there is no reason why the action represented \$Ev should resemble the action represented by vEr in any way. The scheme thus allows for different orderings of the 'same' actions within the same associative structure, but only by suggesting that the same actions are in fact quite different.

2.3 Context-sensitive coding deals, in a similar fashion, with the problem of the interference between sequences in the same associative structure. In this case, associative chains representing the spoken words cat, tack, and act do not interact. This is because allophonic variants of each speech sound have quite different representations in Wickelgren's scheme. For example, the /a/ in "cat" is represented by a token designated kat, whereas the /a/ in "tack" is represented by completely different token, tak. These units have been termed 'wickelphones' (Rumelhart & McClelland, 1986). The use of different tokens for variants of the same phoneme can be used to provide a weak account of coarticulation simply by assuming that each wickelphone is associated with a different articulatory realization. However, Wickelgren's account fails to provide any explanation for the similarity between the same phoneme occurring in different contexts (e.g., the /a/ sounds in cat and tack from the above example).

2.4 In addition to the unsatisfactory use of a token-based representation, Wickelgren's

solution to the problems chaining models face with the representation of repeated items is incomplete. If an action is repeated with identical local context (e.g., in /kankan/), the same wickelphone must be used twice, creating the kind of looping chain shown in Figure 1a. In addition it fails to solve the problem of representing multiple sequences in the same associative structure. For example, all sequences beginning /ka.http://psyche.cs.monash.edu.au/ would begin with the same wickelphone \$ka .To generate the sequence "catalyst" it would be necessary to choose between associative chains radiating from the same starting point ("cat", "camera", "cancel" etc.).

2.5 Some recent neural network models of serial order, in particular Jordan (1986) and related work, overcome some of these problems, while retaining a dynamics dependent on chaining, in that the current output of the network is cued by a learned relationship to some record of its previous responses (or previous internal states, Elman, 1990). These models, made possible by the development of learning algorithms for non-linear mappings, include a static, sequence-specific "plan" input, which helps the networks store different orders of the same items. The use of a history of past outputs, rather than just the last one, as the cue to the next action helps to overcome repetition problems. However, the models require many exposures to sequences to learn them, precluding their use in modeling single trial learning and short-term memory (see below). They also seem to us unlikely to be prone to the kinds of serial order errors discussed below. In addition, the use of chaining still leads to interference between different orders of the same items, constraining learning capacity.

2.6 The attraction of associative chaining lies in its use of well-defined associative learning rules and its avoidance of biologically implausible, computer-based primitives such as "serial buffers" etc. Unfortunately, chaining does not provide a satisfactory basis for the understanding of many aspects of serial learning and recall, whether from long- or short-term memory. There would appear therefore to be room for a theory of serial order which possesses the attractive features of associative chaining (simple learning rules, no intrinsically ordered buffers) while avoiding its limitations.

3. Model 1: Serial Order And Response Competition

"There are indications that, prior to the internal or overt enunciation of the sentence, an aggregate of word units is partially activated or readied" (Lashley, 1951, p. 119).

3.1 It is evident that most associative theories of serial order begin with some prior model for forming an association between two items. The prior model may use simple Hebbian S-R associations, vector convolution (Lewandowsky & Murdoch, 1989), Hopfield learning (Amit, Sagi & Usher, 1990), backpropagation (Jordan, 1986), or some other method. When faced with the problem of extending the basic associative learning model to serial learning, the extension that involves the least additional machinery is chaining. This appears to make it the default option for theorists already wedded to one or another basic model (possibly explaining the remarkable tenacity of the idea). However, this may

not be the most profitable theoretical strategy. All such models basically treat the generation of serial behaviour as little more than iteration over the process of associative recall. The representation of a sequence in memory is thus treated in isolation from any other component of response generation mechanisms. As an alternative approach, we might start with the simpler (and possibly evolutionarily prior) problem of the resolution of response competition in complex environments (i.e., response scheduling under environmental stimulation rather than from memory).

3.2 To make this point clear, suppose an organism is capable of producing two responses r_1 , and r_2 , (e.g., eating and drinking) and these responses are called forth (either innately or due to learning) by stimulus configurations s_1 and s_2 respectively. Suppose now that s_1 and s_2 occur simultaneously, activating r_1 and r_2 in parallel. If the two responses are both valuable but are not such that they can be generated simultaneously, say due to effector limitations, then the organism is faced with the problem either of choosing one response over the other or of *ordering the two responses so that one occurs after the other*, i.e., generating serially ordered behaviour. Either solution requires that one of the activated response tendencies be allowed to control behaviour while the other is somehow "held in abeyance" until the chosen response is completed. For the serial case, the previously withheld response can only be released if the first, "dominant", response is not repeated. This "all or nothing" aspect of animal behaviour has been frequently noted. For instance Hinde (1970, p. 396, cited in Neumann, 1987, p. 377) states: "Undoubtedly the commonest consequence of the simultaneous action of factors for two or more types of behavior is the suppression of all but one of them".

3.3 Response competition due to parallel perceptual processing is commonly observed in experimental situations in which subjects must respond to a target object in the presence of to-be-ignored distractors. The distractors lead to increased error rates and delayed reaction times (Stroop, 1935; Eriksen & Eriksen, 1974). That the effect is indeed due, in part, to response competition is shown in a study by Eriksen, Coles, Morris and O'Hara (1985) using a two-choice reaction time task (see also Coles, Gratton, Bashore, Eriksen & Donchin, 1985). In this study, subjects had to respond to a central letter in the presence of flanking distractor letters, e.g. the "S" in H S H. The distractor letter (H) could appear in the target position on other trials (e.g., S H S), and had a different associated response. Eriksen et al. found that, even when subjects made a correct response to the target letter, the incompatible distractor letter frequently gave rise to its associated response to the point that significant electromyographic activity was detectable in the muscles controlling the relevant effector (the hand, in this case). Reaction times were significantly slower on those trials in which such activity was detected compared to those in which it was not. Subjects in these experiments were required to make only one response on each trial, so the competing response was never released (except in error). It would be revealing to change the design to permit sequential responding, and to measure the degree of activation of upcoming responses.

3.4 Such a scenario is likely to face any organism capable of a degree of parallel processing in its perceptual systems, while limited to largely serial action due to its effector structure (Allport, 1987; Neumann, 1987; Houghton & Tipper, 1994, 1995), and

similar findings are reported for predatory animals faced with more than one prey object (see Ingle, 1972, for an example involving frogs). Thus even simple organisms may be equipped with mechanisms for the selection and serial ordering (scheduling) of responses activated in parallel. This raises the possibility that serial behaviour may be generated from memory by internally activating a set of responses in parallel in such a way that the general "response scheduling" mechanism leads to them being produced serially.

3.5 To develop this idea more concretely, we need first to consider how the response scheduling mechanism might operate. Intuitively, response tendencies can exist with different "strengths", that is, the inclination to perform a particular action can be more or less "pressing". In the simplest case then, given two (or more) competing evoked tendencies, say to take a drink from a glass of beer or take a drag on a cigarette, the stronger will be performed first. The performing of the most pressing action leads to a temporary lessening of its strength ("drive reduction"), leading to the competing action becoming the strongest and hence being produced. This is roughly the kind of mechanism envisaged by Shallice and colleagues (Shallice, 1972; Norman & Shallice, 1986; Cooper, Shallice, & Farrington, 1994) to be involved in the automatic production of routine actions, and given the name "contention scheduling". In this theory, individual response types are represented by response schemata which can be more or less active due to a combination of perceptual inputs (triggering stimuli) and internal motivational inputs (Shallice, 1972). Activated schemata compete by lateral inhibition to become the most active, so that, "No more than one action system may be strongly activated (i.e., become dominant) at any given time." (Shallice, 1972, p. 387). The use of lateral inhibition as the mechanism of conflict resolution essentially means that the initially most active schema (strongest response tendency) will be the one which gains control of effector systems. Note that Shallice's proposals were not developed in the first instance as a theory of serial order but "as a solution to the potential cybernetic problem that an organism has many goals which it needs to achieve at any one time and has only a limited number of effector units available." (Shallice, 1972).

3.6 In the original formulation of this model, Shallice did not discuss how a schema, once dominant, might become de-activated. In this form the model runs the risk of endlessly repeating its dominant response. In addition, the model postulated direct connections from activated response schemata to effector structures, apparently leading to the need for one schema to completely suppress all others in the response competition (to prevent them from sending interfering input to the effectors). As well as being liable to perseverate, this tendency to obliterate competing responses could further cause the model to "forget" other contextually relevant responses. Thus the scenario envisaged above, whereby an animal resolves parallel response competition by sequencing actions could not be easily realized by the architecture. However, these limitations can be overcome by some fairly simple additions. An example is shown in Figure 2.

Figure 2

A mechanism for the resolution of parallel response competition into serial action ("contention scheduling"). Responses activated at L1 compete for control of output at L2. Selected responses inhibit themselves.



3.7 In this model, envisaged as a neural network, one layer of nodes (L1) corresponds to Shallice's response schemata. The activation of a response is a continuous variable in some range, represented by the activation value of a given node. Activating inputs (from whatever source) arrive at response nodes in parallel, and more than one node may be simultaneously active. Instead of proposing that conflict resolution must take place at L1, this function is devolved to another layer of units (L2 - the "competitive filter", Houghton, 1990). In the simplest case L1 nodes can activate L2 nodes in a one-to-one fashion. The response activation in L1 is thus copied to L2. It is proposed that the severe lateral inhibitory (competitive) interactions envisaged by Shallice take place at L2, so that the initially most active node suppresses the rest. This scheme means that response selection can take place at L2, without the need to completely suppress the representation of other potentially important responses, which can remain activated at L1. The need to suppress the currently dominant response after completion (to prevent perseveration) suggests the use of some form of inhibitory feedback to L1. In principle this could be quite a complex process, depending on the internal complexity of the response. In the simplest case however, we can imagine that, once selected for output at L2, the activation of the response at L1 is no longer needed. Thus a simple one-to-one inhibitory feedback loop from L2 to L1 will cause the selected response to inhibit itself (Figure 2). Once this is done, the remaining activated responses at L1 can compete to be produced next. It is easy to see that if a set of responses are activated in parallel at L1, but with a "gradient" of activations over them representing response strength, then this mechanism, though entirely parallel in itself, will sequentially select responses in the order dictated by their degree of activation. In other words, serial order can be an emergent property of a parallel mechanism dedicated to resolving response competition.

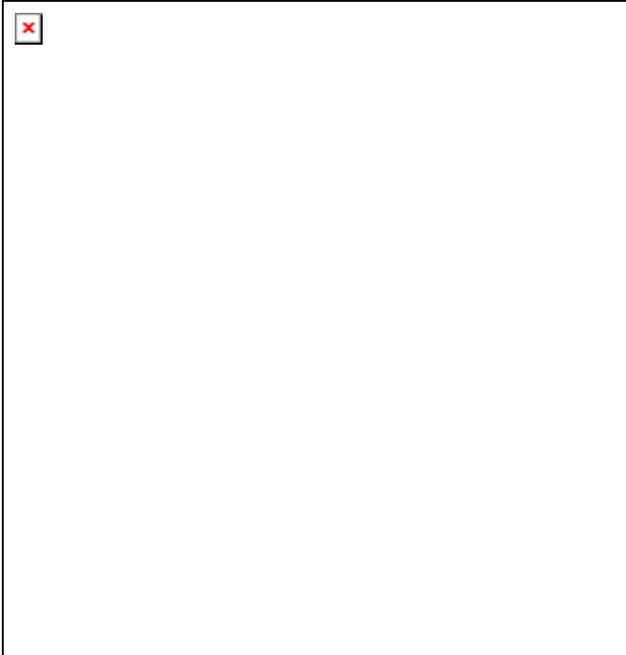
3.8 If a mechanism with the basic characteristics described above is in place to enable an organism to order its actions in terms of some simple internal measure of response strength, then the organism could produce serially ordered behaviour from memory simply by being able to activate all the responses in some sequence in parallel, but with

an activation gradient over the responses, such that the sooner the response is to be produced the more active it is. Specific models with this basic character have been proposed for a number of serially ordered behaviours (Estes, 1972; Grossberg, 1978; MacKay, 1987; Rumelhart & Norman, 1982), though with by far the greatest emphasis on one or another form of linguistic behaviour. A good example is provided by the Rumelhart and Norman (R&N) model of typing. In this model, R&N were particularly interested in the form of typing errors and how they could constrain models. Error data (e.g., MacNeilage, 1964; MacKay, 1970, 1972; Norman, 1981; Reason, 1984) have played an especially important role in the development of the serial order models described below. For instance, many typing errors are transposition errors of the form "trap" ->"tarp". Transposition errors are common in many serial order tasks (e.g., immediate verbal recall tasks) and are highly problematic for most conventional models of serial order. Consider what has to happen to produce the error "trap" ->"tarp". First, at the point at which the "r" should be produced two things occur: the "r" is not produced, but the "a", which should occur later, is produced instead. This provides evidence that upcoming responses are already active before the point at which they are to be produced. Second, at the point at which the "a" should be produced two further things happen: the "a" is not produced in its correct position (i.e., it is not repeated), and the "r" which was omitted at position 2 is now produced. These events provide evidence that the "a" response produced at position 2 has been suppressed, preventing its occurrence at its appropriate position. If this were not so, this would produce an error such as "taap" or "taarp", forms which are rarely, if ever, found. The idea of suppression is further supported by the appearance of the "r" in position 3, indicating that, not having been produced at position 2, it has remained active.

3.9 The Rumelhart & Norman model is outlined in Figure 3. The model is hierarchical, in that specific sequences are represented as "chunks", i.e., sets of individual responses bound together by connections to a higher-level node which spans the chunk. The chunks in the model correspond to words (or parts of words). When a word is to be produced nodes representing the letters in the word are equally activated in parallel by word-to-letter connections. Letter nodes in a chunk have lateral inhibitory connections between them such that each one is inhibited by those nodes representing letters which are to precede it (a scheme suggested by Estes, 1972). The first letter therefore receives no inhibition and later letters receive progressively more. Thus the net excitation (excitation from the word node minus inhibition from letter nodes) received by a letter node decreases the later in the word it is to be produced. This induces an activation gradient over the letter nodes. In combination with a select-and-inhibit mechanism of the type described above (Figure 2), this parallel activation generates serial output.

Figure 3

Schematic diagram of the Rumelhart and Norman (1982) Typing model.



3.10 The R&N model produces errors by the addition of noise to letter node activations. Transposition errors occur when the wrong letter node becomes the most active at the wrong time. Given the activation gradient, this is most likely to be the letter to be produced at the following position - thus most transpositions involve adjacent letters, as is found in the human data. After the wrong response has been produced it is automatically inhibited, preventing it being repeated at its correct position. The omitted response remains active however, allowing it to win the output competition at the next position. The model thus produces these errors quite naturally, in an intuitively satisfying way. R&N also model other a priori puzzling error types, such as "doubling shift" errors of the form "screen" -> "scrrer" (discussed in more detail below). In addition, the parallel response activation in the model is independently motivated by its use in modeling co-articulation effects in typing, in which the hand configuration adopted by the typist while making one response is affected by the location (on the keyboard) of upcoming key presses.

3.11 The R&N model illustrates the explanatory value of a serial order model based on the resolution of parallel response competition. What, though, is its status as model of memory for serial order? This question involves the internal representations (connection patterns, in this case) by which it generates the necessary activation gradient over the response set. This has two components. One component, the word-to-letter connections, provides equal excitation to all letters in the word. As this is the only activation letter nodes receive, this input specifies item information, i.e., what letters are in the word. The other component, the lateral inhibitory connections, specifies order information (the activation gradient). This latter component is problematic, for reasons similar to the problems with chaining models discussed above. To be plausible, the model must represent the spelling of all words (or, at least, plausible subword chunks) with the same

set of letter nodes. Many words obviously contain the same letters in different orders (e.g., trap, part, rapt). The lateral inhibitory connection pattern needed to specify one of these orders is obviously different from that required for the others. If all the patterns are simultaneously present in memory then they will clearly interfere with each other. Indeed, since the word nodes for the above three words all activate the same letters to the same degree, the lateral inhibitory pattern due to the representation of all three words in memory will lead to exactly the same letter node activation pattern whichever word node is active. Other problems arise for instance in words containing repeats, such as "prop", in which the "p" is first and last, and hence must be simultaneously the least and most inhibited letter. In these cases, R&N have to parse the sequence by making divisions on the occurrence of (non-immediate) repeats. Although such a chunking scheme has some attractions (see Keele & Jennings, 1992, for a similar proposal), it can lead to parsings of words which are not the most intuitively appealing, e.g., trot -> (tro)(t), leaning -> (leani)(ng), disastrous -> (disa)(strou)(s), nonetheless -> (no)(neth)(el)(ess).

3.12 Rosenbaum, in his book on motor control (1991; p. 285), says of the Rumelhart and Norman model that "It represents an important advance in the modeling of human motor control and should serve as a useful starting point for future research". Below we develop models which show similar behavioural characteristics, but which do so on the basis of learned representations of serial order that do not have the problems discussed above.

4. Model 2: Response Competition Under Internal Modulation

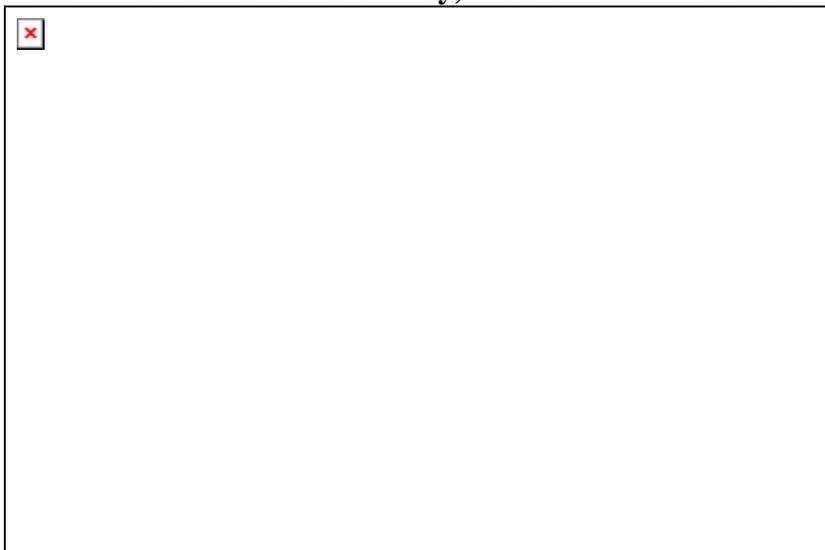
"My principal thesis...will be that the input is never into a quiescent or static system, but always into a system which is already actively excited and organized. In the intact organism, behavior is the result of interaction of this background of excitation with input from any designated stimulus. Only when we can state the general characteristics of this background of excitation, can we understand the effects of a given input" (Lashley, 1951, p. 112).

4.1 Recent work on serial order involving the modulation of parallel response competition has developed learning algorithms for these models which produce memory representations that do not suffer from the difficulties facing the Rumelhart and Norman model (Burgess & Hitch, 1992; Houghton, 1990). Following Houghton (1990), we will henceforth refer to models based on such principles as "Competitive Queueing" (CQ) models. This name reflects the idea that the activated responses in such models are "queueing" for "service" (i.e., output), but without forming an ordered line, such as might form at a ticket office. A competitive queue is more analogous to the situation at a crowded bar with only one bartender. Customers are still served one at a time (serial order), but no ordered structure exists. Service depends instead on success in the competition to attract the bar staff's attention.

4.1 In recent CQ work, the storage of sequence information in connections (excitatory or

inhibitory) between response elements is avoided. Instead, it is contained in connections to the response elements from nodes at a higher level. In the Rumelhart & Norman model, such connections (from the word to the letter level) all have the same strength and thus contain only item information (what letters are in the word). It is possible, however, for these connections to contain order information if their strengths are allowed to vary. For instance, a word node might have a stronger connections to a letter the earlier it appears in the word. Activation of the word node would then activate all the letters in the word but to a degree dependent on the letters' target positions. Response selection mechanisms of the type discussed above would lead to serial output in a similar manner to that achieved by the Rumelhart and Norman model. Note that in this scheme there is no anagram problem, as there are no sequence-specific lateral connections. Figure 4 shows the representation of the words rat, art and tar using such a scheme. Activation of one order of the three letters is not affected by the ability to activate any other order.

Figure 4
A hierarchical coding of response gradient. Anagrams RAT, ART and TAR can be encoded purely by word-to-letter connections, avoiding any cross-talk. All connections shown are excitatory, darker connections indicate stronger weights.



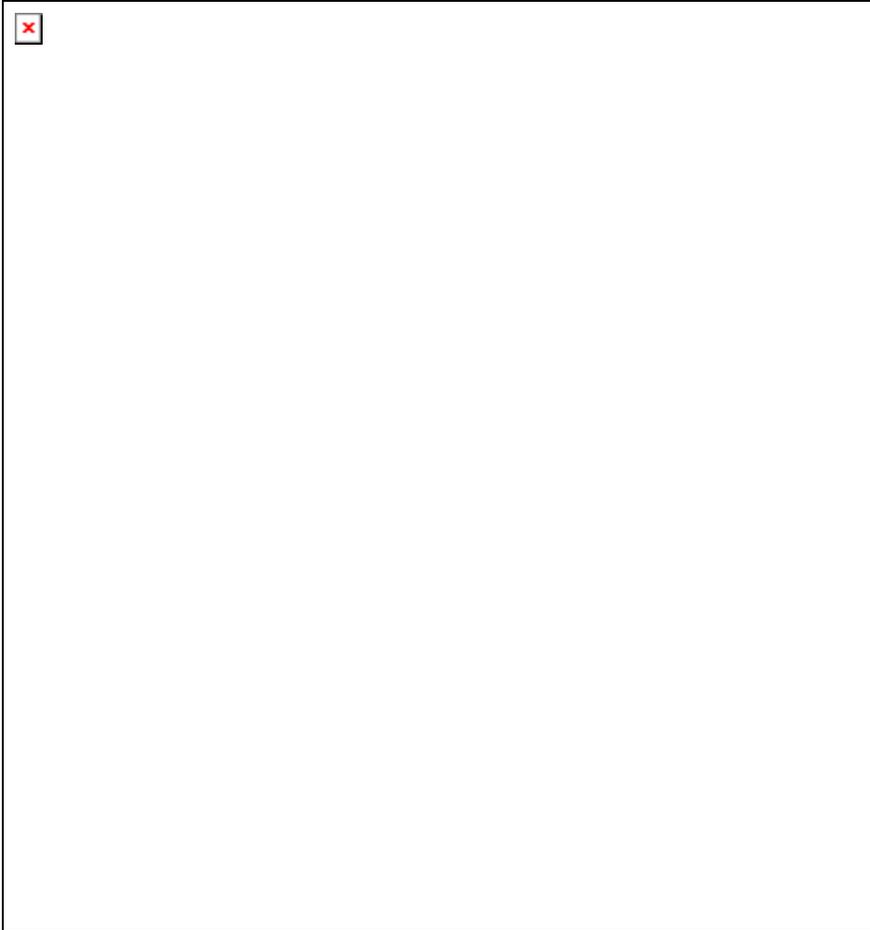
4.2 This scheme still faces problems however. One the one hand, the suppression of responses after production makes difficult the storage of a sequence such as prop, as the letter p has to be produced twice. The arrangement shown in Figure 4 would simply produce pro. Conversely, there is also the problem of the undesired reactivation of previously executed responses. Though suppressed, they will continue to receive activation from the word node while it is still on. If suppression decays (as it surely must), then this top-down activation could be sufficient to reactivate items produced early in the sequence (which receive the strongest inputs). This problem would become worse the longer the sequence to be stored.

4.3 Recent CQ models incorporate an important development aimed at solving both of

the above problems (Houghton, 1990). The basic idea is to use a dynamical representation of serial position at the level above the items to be sequenced. We may refer to this level in the general case as the "sequence" or "control" level. In the Rumelhart and Norman model, no positional information (i.e., information as to where one is in the sequence) is available from a word node, as this comes on at the beginning of the sequence and then stays on without varying. Sampling of the state of a word node tells one whether the sequence that node stands for is being produced, but not what point in its execution has been reached. Positional information can be incorporated into these models by abandoning the assumption that activation at the sequence ("word") level should be (i) static, and (ii) unidimensional. That is to say, one permits a sequence to be controlled by the activation of a set of "sequence nodes" whose state of activation varies in some regular way during learning and execution of a sequence (Burgess & Hitch, 1992). Formally, the activation of the sequence level, rather than being a scalar constant, becomes a time-varying vector, which we may refer to for convenience as the "control signal". (Figure 5). This vector can be used to implicitly encode positional information. During learning, different states of the signal become associated with different response states, according to simple learning rules (Houghton, 1990). This adds an endogenously dynamic element to these models, so that, in Lashley's words, the input is "never into a quiescent or static system".

Figure 5

Recall in Competitive Queuing (CQ) models employing a time-varying "Control Signal" (CS). The shaded vertical bars in the graphs represent the degree of activation of nodes representing the letters P, R, O during recall of the sequence prop. At each time (t3d1-4), the most active response is selected and subsequently inhibited. the shaded horizontal bar represents the control signal, conceived of as a time-varying vector. The shading of the bar represents the "degree of activation" (magnitude) of components of the control signal vector. Darker shading represents more activation. Different responses are associated with different states of the control signal, and repeated items may be associated with more than one state. This is shown by the arrowed lines from the CS to the letters. The crucial factor in generating the characteristic CQ activation gradient (responses being more active the sooner they are to be produced) is that the state of the control signal changes smoothly and monotonically, being more similar to itself at closer positions in time.



4.4 With this mechanism, a sequence such as PROP can be stored because the two occurrences of the letter P are associated with different states of the control signal. This is illustrated in Figure 5. Here the control signal is represented schematically by the shaded horizontal bar, the degree of shading representing degree of activation, and the change in shading representing changing activation. The arrows from the control signal to the letters represent associations between particular components of the signal and those letters. The P node is activated twice, the first occurrence activated by the "start state" of the signal, and the second by the signal moving towards its end state. The same mechanism solves the problem of the undesired reactivation of responses. If the control signal changes quickly enough, or a response node discriminates sufficiently between successive states of the signal, then a suppressed response will not be reactivated because the evolving state of the control signal will soon cease to be strongly associated with a response once its target "position" has passed. For instance, in Figure 5, the end state of the control signal, which reactivates the P, is not strongly associated with the R, which remains suppressed. If the target word were, say, PROD, then the P would not be associated with the end state and would not be reactivated.

4.5 An important constraint on the form of the control signal is that it should be temporally correlated (more similar to itself at nearer points in time) in order to produce

the characteristic CQ activation gradient, whereby responses are more active the sooner they are to be produced. If the CS has this property, then any given state will partially activate responses associated with similar states, and these responses will be ones that occurred at similar times or positions. This is indicated in Figure 5 where the shaded bar changes gradually.

4.6 How complex does the control signal need to be? In the absence of additional specific constraints, it seems desirable to investigate the properties of the simplest that are likely to work. Work by Houghton and colleagues (Houghton, 1990, Houghton et al., 1994) has shown that a control signal generated by two nodes, one starting with high activation and then falling (a "start" node), and the other starting with low activation and then increasing (an "end" node), can encode sequences of lengths up to around seven or eight items, including ones with repeated items. Houghton et al. (1994) use this form of control signal in a model of lexical spelling. The signal is smoothly correlated in time, generating an activation gradient similar to that of the Rumelhart and Norman model, but without the use of sequence specific lateral connections. These models show that it is not necessary to use a "discrete" positional representation, i.e., one in which specific nodes represent specific positions, though such a representation is compatible with the general approach. Burgess and Hitch (1992) use a more powerful "distributed" representation of position (referred to as the context) in which individual nodes represent more than one position, and each position is represented by more than one node. This model is especially effective for single-trial serial order learning, but the complexity of control signal begs the question of its origin.

4.7 The idea that serial behaviour might depend on the existence of such internal dynamics relates to similar ideas in models of time perception, in which the endogenous activation (the internal "clock") is typically generated by oscillators (Church & Broadbent, 1990; Treisman, Cook, Naish, & McCrone, 1994). Similarly, motor sequencing in many species has been found to depend on the generation of repeating patterns of activity by groups of neurons known as "central pattern generators" (Pearson, 1993). The neurons in the pattern generators are distinct from the motoneurons controlling individual responses. Recent work by Burgess, Hitch and colleagues has suggested that the internal signals required for learning and recall in their short-term memory model might be composed of oscillators entrainable to the rhythmic characteristics of the input sequence (Hitch, Burgess, Towse & Culpin, 1995; Hitch, Burgess, Shapiro, Culpin & Malloch, 1995).

4.8 Such mechanisms have been applied in a number of a domains including speech production (Houghton, 1990; Hartley & Houghton, in press), auditory-verbal short-term memory (Burgess & Hitch, 1992; Burgess, 1995; Glasspool, 1995; Hitch, Burgess, Towse & Culpin, 1995; Houghton, Hartley & Glasspool, in press), speech errors in immediate nonword recall (Hartley & Houghton, in press), and serial order errors in spelling (Houghton, Glasspool & Shallice, 1994; Shallice, Glasspool & Houghton, in press). Full review of this work is beyond the scope of the current paper. Instead we will

concentrate on two issues covered in this work which provide important sources of constraint on serial order models. The first involves the problem of single-trial serial learning, the second the importance of error data in recall.

Serial Order and Short-Term Memory

4.9 Experimental studies of short-term memory have frequently employed serial recall tasks, where subjects are required to reproduce, in correct order, an unfamiliar sequence of familiar items, e.g., digits. If the lists are not too long, the most common type of error in such studies is the misordering of the items in the list. It is thus the novelty of the sequence, rather than its content, which seems to test short-term memory. If a well-known sequence, such as the days of the week, is presented then such errors will be much less likely. The crucial factor in such studies is that the sequence information must be encoded on-line, in a single trial. This is especially important in serial recall of nonwords (e.g., Treiman & Danis, 1988), in which not only the order of the items must be learned on-line, but also the order of the phonological units making up the nonwords. Although in the past nonword recall has frequently been cited as a prime example of experimental tasks which are ecologically bizarre, recent discoveries have shown nonword repetition ability to be causally related to capacity for long term phonological learning, an important component of vocabulary acquisition and language learning generally (see Gathercole & Baddeley, 1989, 1993; Gathercole & Martin, in press; Service, 1992; Papagno & Vallar, 1995; Baddeley, Gathercole, Bishop & Papagno, submitted).

4.10 Such data show that the ability to rapidly encode serial order is of considerable importance for human development (at least in speech, though we suspect the same may apply to the imitation of movement more generally). In addition to its ecological significance, rapid learning offers a powerful constraint on theories of serial learning. For instance, the recurrent sequential networks developed by Jordan (1986) and Elman (1990) require the use of iterative, supervised learning procedures, and hence are unable to effect single-trial, unsupervised, learning. CQ models generally learn using unsupervised "Hebbian" learning, i.e., weights between nodes at the sequence and item levels are adjusted as a function of their co-activation. Such learning can take place on-line. Thus associative learning models for short term memory can be developed which do not rely on inter-item chaining.

4.11 A number of models of STM with this character have been developed (Burgess & Hitch, 1992; Burgess, 1995; Glasspool, 1995; Grossberg, 1978; Hartley & Houghton, in press; Henson et al., in press). Although differing in various respects, all these models learn rapidly without forming inter-item links, and achieve sequencing through parallel response activation and competition for output. They are thus all prone to error types such as immediate transpositions and others which, while commonplace in human data, cause serious problems for non-queueing models (Henson et al., in press). Other characteristics of serial recall such as bowed serial position curves (primacy and recency effects) have also been shown (Burgess, 1995).

4.12 These models have been specifically applied to auditory-verbal STM, and the

question arises whether they might apply to the rapid learning of other kinds of action sequence. It is important to note that verbal recall may have a number of idiosyncratic properties related to the nature of the to-be-retained materials. For instance, verbal recall shows word length effects (Baddeley, Thompson & Buchanan, 1975; Cowan, Day, Saults, Keller, Johnson & Flores, 1992), effects of phonetic confusability (Conrad & Hull, 1964; Baddeley, 1968), and effects of the lexical status (word/nonword) of the list items (Hulme, Maughn & Brown, 1991; Treiman & Danis, 1988). However, in the models mentioned above, such effects are typically due to factors other than the competitive queueing dynamics underlying serial ordering. As such they can be treated separately from the basic issue of ordering, and the production of movement sequences could be studied for the presence of such basic features as the preponderance of order errors and the serial position curve. A study by Wilberg (1990) provides some evidence that memory for other kinds of action sequences may indeed depend on similar principles to auditory-verbal memory. Although Wilberg's study used free-recall of movement lists (as opposed to serial recall), he found strong evidence for typical order effects in free recall, including primacy and recency effects. Wilberg concludes that his results "suggest that memory for movement and memory for words are not substantially different." Further studies of this type, particularly using serial recall and involving detailed analysis of errors, would be highly instructive for the issue of the general foundations of motor sequencing.

Errors in Serial Recall: The Problem of Repetition

4.13 As has been repeatedly emphasized, the analysis of error data has been especially important in motivating the type of model discussed above (MacKay, 1970, 1972, 1987). We further illustrate this point with a particularly puzzling error which Lashly noted: misplaced repetitions.

4.14 The use of response inhibition in CQ models is necessary for them to function properly and is central to the account they provide of transposition data. However, it leads to an obvious problem: How can an action be immediately repeated, e.g., typing the letter "p" in the word "supper" ? If the action is suppressed after being produced, then, in a CQ context, the next most active response will be generated. Thus "supper"

would be typed "super". This problem can be overcome if one postulates that response repetition is a special "mode", which is entered into only occasionally, i.e., the default assumption in behaviour is that successive actions will be different from each other, and that perseveration must be avoided. A similar assumption appears to be built into the movement of attention, leading to an "inhibition of return" (IOR) effect (Posner & Cohen, 1984), whereby attention is slower in returning to a recently attended location than in moving to a new one.

4.15 This basic assumption is clearly part of the competitive queueing architecture. The hypothesized "repetition mode" acts in some way to prevent the usual response inhibition from taking place, allowing a given response to be repeated while the mode is active. Repeated letters can be produced if the mode is invoked at the appropriate point in

production, and only remains briefly active. Essentially this proposal is made by Rumelhart & Norman (1982) in their typing model. Any letter to be doubled in a given word is associated with a "doubling schema", which ideally becomes active when the letter wins the output competition. This temporarily disables the usual inhibitory feedback, allowing the letter to be repeated (in the absence of inhibition, it remains the most active response).

4.16 This may appear a rather ad hoc solution to a problem that one would prefer not to have in the first place. However, the doubling schema idea has empirical consequences. Errors occur in the R&N model due to noise in activation levels. The doubling schema, like letter nodes generally, is itself subject to noise, and it will occasionally become active slightly before or after the appropriate point. This leads to the wrong letter, generally a flanker of the target letter, being doubled, e.g., "supper" -> "suuper". It turns out that such errors are commonplace in typing (readers will undoubtedly find them in their own typing), and they generally involve letters adjacent to the target letter. Similar problems are occasionally found in handwriting, but are less common (possibly due to the generally slower pace of handwriting). However, they have been found to occur in subjects with an acquired neurological disorder known as graphemic buffer disorder (Caramazza & Miceli, 1990), who show specific and comparable impairments in the spelling of both words and nonwords. Houghton, Glasspool & Shallice (1994) model lexical spelling with a learning CQ model based on that of Houghton (1990), but incorporating Rumelhart & Norman's doubling schema to allow the model to learn words such as "supper". The graphemic buffer disorder is modeled by the addition of debilitating amounts of noise to the letter nodes activated when a word is to be spelt. Addition of noise to the doubling schema leads to spelling errors involving misplaced double letters.

4.17 Thus the implicit postulate of CQ models that behaviour has a built in tendency not to repeat itself leads to the requirement for a specific behavioural "mode" when repetition is required - in this mode the mechanisms which normally keep behaviour "moving forward" are suppressed. When sequences with repeats are learned, the point at which this mode must be entered has to be encoded. Errors in retrieval of this point during recall can lead to the wrong action being repeated. An additional prediction derivable from this idea concerns leaving the repetition mode. Clearly, if the mode is not turned off, then the dominant action will continue to be repeated. It has been found in handwriting and typing that letters which should be doubled are sometimes tripled (Ellis, 1979), indicating that the repetition mode has not been turned off sufficiently quickly. Ellis, Young and Flude (1987) report (handwritten) spelling errors made by an acquired dysgraphic patient, the majority of whose addition errors involved producing too many copies of doubled letters, e.g., ladder -> laddder, chilly -> chilllly. Venneri, Cubelli, and Caffarra (1994) report a similar case of the handwriting of an Italian dysgraphic patient who only produced letter perseverations in words containing a doubled letter. In most cases the perseveration consisted of tripling a doubled letter, though repetitions up to 6 letters long are reported. Few perseverations of single letters occurred, and all but one were found in words which contained a doubled letter elsewhere in the word (e.g., parallelo -> parallello). Such data

support the idea that doubling involves a "repetition" mode, which results in perseveration if not terminated.

5. Serial Order and the Origins of Grammar

"This is the essential problem of serial order: the existence of generalized schemata of action which determine the sequence of specific acts, acts which in themselves or in their associations seem to have no temporal valence" (Lashley, 1951, p. 122).

5.1 In all the models and data considered so far, the items being sequenced are individual responses which are directly activated by input from the "sequence level", be it a steady state input such as in the Rumelhart and Norman model, or a time-varying one. If more than one sequence is stored, then each has its own dedicated sequence node(s) activating the appropriate responses (Figure 4). But, as Lashley emphasized, many individual action sequences appear to be exemplars of a more general "schema for action". This brings us to the issue of grammar (or syntax), the representation of generalized sequential patterns, whose individual components can vary. Thus for instance, in English the word order of many simple noun phrases (NP) may be specified by the phrase structure rule: NP → det adj noun, where det 3D determiner, adj 3D adjective, → 3D "is realized as", and the left-to-right order of the symbols following the arrow represents serial order. The crucial difference between such a representation and anything considered above is that the ordered items (det etc.) are variables, rather than specific responses such as individual words. These variables range over particular classes of word (or "lexical item"); for instance determiner may be realized as "the", "a", etc., adjective as "big", "small", etc., noun as "girl", "boy" etc. The rule given above can be used to generate or describe numerous sequences or words by instantiating each variable by a word from the appropriate class, e.g., the small boy, the big meeting, a dainty biscuit etc.

5.2 It seems impossible to account for the productivity of language use without recourse to some form of "schema of order" which is not defined (solely) in terms of specific words. Nonetheless, it has been argued (e.g., Ellis, submitted) that knowledge of language cannot be properly captured solely in terms of such abstract schemata either, and that native speakers routinely employ numerous preconstructed phrases (idioms etc.) in which specific words are specified, for instance, "How are you?", "A stitch in time saves nine", "You could have knocked me down with a feather". Other expressions contain mixtures of words and grammatical variables. For instance, Ellis (submitted) gives the example "NP be-tense sorry to keep-tense you waiting", where the italicized items are variables. The schema can be realized as "I'm sorry to have kept you waiting", "Mr. Brown was sorry to keep you waiting", etc. Our impression is that "idiomatic" command of a language may depend almost as much on knowledge of such formulae and the conditions of their use as it does on the kind of abstract, generative, knowledge studied in linguistics.

5.3 Such examples suggest a "productivity" continuum, with clichés, proverbs and other formulae at one end, and fully "creative" language use (e.g., poetry) at the other. Most ordinary language use appears to fall somewhere in between, suggesting a cost-benefit

trade-off. The benefit of encoding specific words is presumably that retrieval of a prespecified word or phrase is a simpler (faster) operation than choosing a grammatical schema and then filling it out through lexical selection. The benefit of using variables is that the same schema can be used in different situations, with the variables instantiated appropriately. The alternative of only storing fully instantiated word sequences would impose heavy memory costs (each sequence being stored separately), and lead to a loss of adaptability. Thus one can argue that "knowledge of language" trades off speed/flexibility benefits against space/time costs. This perspective requires that language learning involves the retention in memory of verbatim sequences of words, for the acquisition of idioms and formulae, and suggests that the further development of abstract schemata is also based on this learning. Long term retention of formulae requires the short term retention of examples of them, and hence acquisition of "grammar" may depend on the integrity of verbal short-term memory (as argued for instance by Speidel, 1993; Baddeley et al., submitted; Ellis, submitted). This reinforces the point made earlier of the importance of the development of models for the rapid learning of serial order.

5.4 Is there any evidence to indicate that the kinds of sequencing principles we have discussed so far can be profitably extended to domains involving grammatical (variable-based) ordering? Evidence from lexical level speech errors indicates that they can (Cutler, 1982; Dell, 1986). For instance, just as letter transpositions occur in typing, whole word transpositions occur in speech. For instance Garrett (1976) gives examples such as: "... but a beach on the bikini is all right" ("beach" and "bikini" exchanged); and "It waits to pay" ("pay" and "wait" exchanged, with morphological normalization). A priori possible errors such as "It waits pay to", or "It tos wait pay" are never found. Similar conclusions may be drawn from these cases as were drawn from the discussion of exchange errors in typing, viz. words are active before being produced, appear to be inhibited following production (not reappearing at the appropriate position), and remain active if they have been omitted. The crucial additional factor in these examples is that what exchanges with what is constrained by the grammatical form of the intended utterance. In the first of the above examples, two nouns are exchanged, in the second two verbs. The result of this is that the intended grammatical form of the utterance is maintained even though the order of words is altered. This is particularly clear in the second example, "It pays to wait" -> "It waits to pay", where the misplaced verbs have adopted the appropriate morphological forms.

5.5 The addition of these grammatical constraints appears to have the consequence that all concurrently active words do not compete on an equal footing at any given position. Rather the competition is largely confined to those words which grammatically match the current target word (MacKay, 1987). At a "noun" position only nouns compete, at a "verb" position only verbs, and so on. This indicates that the grammatical variables "noun", "verb" etc., are entities over which serial order is defined. They may be thought of loosely as being associated with "slots" in an utterance, with the slots being activated in sequence. As each slot becomes active, it selects a lexical item from those currently active to fill it. The lexical item must be of the appropriate type for the slot, but if more than one such item is active then they compete to occupy the slot, possibly on the basis of their activation level.

5.6 Compare this situation with the typical verbal STM task, in which there is no grammatical structure in the stimulus lists and all stimulus items tend to be of the same sort, letter names, digits, nouns or whatever. No position in the list therefore has any distinguishing features associated with it. In this case, CQ models predict that the output competition will be solely based on item activation, and that items will be more active the sooner they are to be produced (Burgess, 1995). The strongest competitor to a given item at a given position will therefore be its intended successor, and this is thus the item most likely to be involved in a transposition error (Henson et al., in press). If we add in constraints of a grammatical nature, then the strongest competitor to, say, a given noun will be the noun planned to appear in the next noun slot. In a sense, grammatically constrained errors of the type illustrated still involve "immediate" neighbours, if neighbours are defined to be of the same grammatical class.

5.7 Models of the type discussed in this paper, involving parallel retrieval, response competition, post-output suppression etc., have been proposed in the domain of syntax in utterance production (e.g., Dell & Reich, 1981; Mackay, 1987). For instance MacKay's diagrammatic model postulates the existence of grammar nodes which, when activated, equally activate a set of word class nodes in parallel, e.g., a particular "noun phrase" node might activate ("prime" in MacKay's terminology <1>) nodes det, adj, and noun. Like Rumelhart and Norman, MacKay uses the Estes (1972) solution to generate an activation gradient over these word class nodes - so the det node inhibits adj and noun, and adj inhibits noun. Firing of word class nodes follows the "most primed wins" principle, with nodes being inhibited after firing. Word class nodes are connected to nodes representing all words in the class, so that, for instance, the adj node is connected to all adjectives. Its firing alone therefore does not pick out any particular adjective to produce. This is achieved by semantic input. When a noun phrase having the target structure (e.g., the black dog) is to be produced, it is postulated that, in parallel with the activation of the word class nodes, semantic nodes (representing the meaning of the noun phrase) directly activate words which express that meaning. This typically leads to more than one word being active (or "primed"), without, in itself, specifying their order. Utterance production proceeds by the combination of the semantic activation of specific lexical items (content) and nonspecific input from the sequential firing of the lexical category nodes (structure).

5.8 Natural language syntax shows considerable complexity, and is clearly not exhausted by the specification of word order. Development of schema based sequencing within the present framework might therefore benefit from considering somewhat simpler examples. In our own work (Hartley & Houghton, in press), we have looked at another example of "grammatical" constraint in language production, the order of phonemes in syllables. Many linguistic studies have shown that syllable structure is universally constrained according to a number of principles, including the "sonority principle" (Selkirk, 1984), and the "resolvability principle" (Greenberg, 1978; Hjelmslev, 1936. See Houghton, Hartly, & Glasspool, in press, for discussion). In addition, particular languages show idiosyncratic constraints. For instance, German allows the syllable initial consonant cluster /shl/, whereas English does not, even though it permits both phonemes to occur in those positions otherwise (cf. shrink, sleep). Taken together, such constraints have the

effect that only a small proportion of the sequences definable over the phonemes of a language actually occurs. For instance, Houghton et al., (in press) estimate that of the a priori possible English initial consonant clusters (including singletons), only about 0.43% actually occur (this estimate is based on figures from Greenberg, 1975, and excludes clusters containing a repeated phoneme).

5.9 Phonological speech errors involving phonemes, such as "barn door" -> "darn boor" (Baars, Motley, & Mackay, 1975), conform strongly to these syllabic constraints. The universal constraints are effectively never violated in phonemic speech errors, and the syllabic structure of planned syllables tends to be maintained, even though individual phonemes may be substituted (Dell, 1986; Ellis, 1980; Treiman & Danis, 1988). In the area of spontaneous speech production, studies of such error data have led to the development of models such as those of Shattuck-Hufnagel (1979), and Dell (1986, 1988). Although these models differ in many important respects, they share the central features that (i) syllables are a fundamental unit of speech planning, and (ii) the structure and content of syllables are separately represented. The content of a syllable can be represented as a set of phonemes. The structure may be represented, as suggested above for grammatical structure, by a set of slots, each of which can only be filled by a subset of phonemes. In a typical syllable, the initial and final slots will be for consonants and the middle slot(s) for vowels. According to the sonority principle, the consonant slots nearer to the vowel are occupied by more sonorous consonants, such as liquids and nasals.

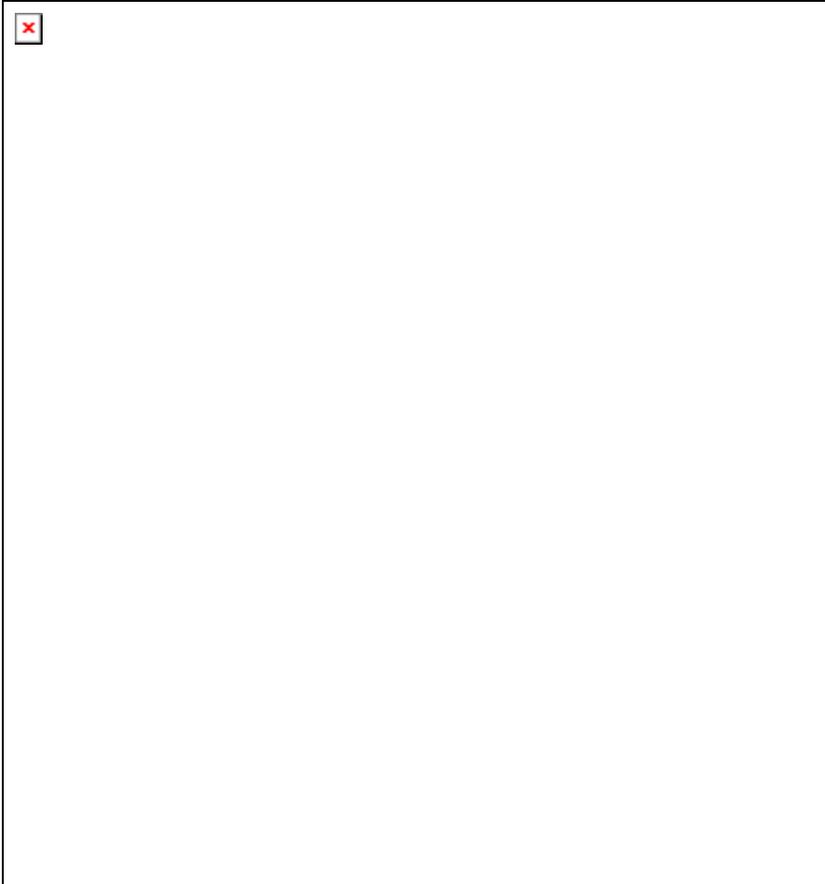
5.10 Hartley & Houghton (in press) develop a model of short-term memory for nonwords based on this idea (combined with the general principles of competitive queueing). Nonword recall was chosen for a number of reasons. All the words a speaker knows were effectively nonwords on first hearing them, and repetition and rehearsal require single-trial phonological learning. As noted, recent studies of phonological short-term memory have shown the importance of such abilities for language acquisition (see Baddeley et al., submitted, for review). Work by Gathercole and colleagues has led them to the conclusion that nonword recall is a more sensitive test of phonological STM than recall of word lists. For instance Gathercole and Baddeley (1993; p. 48) state that, "[P]erformance on immediate memory tasks can reflect the contribution of long-term memory knowledge as well as short-term memory processes...We therefore expect to gain a more sensitive measure of phonological memory skills by using memory items for which there are no long-term lexical representations, because subjects will be less able to use lexical knowledge to supplement phonological short-term memory." In addition, work by Treiman and colleagues (Treiman & Danis, 1988; Treiman, in press) has shown that phonological errors in the recall of nonwords are much more frequent than for words, but are constrained in the same way by principles of syllable structure. Hartley and Houghton (in press) propose that the capacity for single-trial phonological encoding exploits existing knowledge of such structure.

5.11 In the Hartley and Houghton model (Figure 6), incoming verbal stimuli are parsed into syllables. When a new syllable is to be learned an onset/rhyme node pair is activated. As each phoneme in the syllable arrives it activates a different slot in a generalized syllable "template", using long-term associative knowledge. Connections from the

activated onset/rhyme nodes to the phoneme and template nodes are strengthened by a Hebbian weight change rule. The connections to the phoneme nodes learn what phonemes occur in the syllable (phonemic content), while the connections to the syllable template learn which positions are used (syllabic structure). Figure 6 shows the representation of the syllable /rat/. Recall of the syllable involves both recovery of its constituent phonemes and the serial reactivation of the syllable slots activated by those phonemes during learning. Phoneme nodes therefore receive input from both onset/rhyme nodes and the syllable template. As each syllable "slot" becomes active, phonemes compete for selection for output. However, the competition is biased strongly in favour of those phonemes which "fit" the currently active syllable slot. If the model is recalling a series of syllables, the strongest competitors for a given slot will be phonemes from upcoming syllables (already active, due to general CQ principles) which occur in the same respective position. Thus errors tend to involve movement of a phoneme from one syllable to the same position in another. The output of the model is tested in detail against data from short term memory experiments (Treiman & Danis, 1988; Treiman, in press), and the nonword repetition of both children (Gathercole et al., 1991) and neurologically impaired subjects (Bisiacchi et al., 1989).

Figure 6

Phonological representation of a syllable (/rat/) in the Hartley and Houghton (1995) model. Not all connections or nodes are shown. Strings of input phonemes are divided up into syllables (syllable group). Syllables are represented in terms of their phonemic content (phoneme group) and the "slots" they use in a generalised syllable template. Syllable group nodes are composed of pairs of onset and rhyme nodes. The solid lines represent temporary weights, formed during rapid learning (short-term memory). The dashed lines are permanent connections (long-term memory). syllable structure and content are separately represented, but interact during recall. Key: On 3d onset, Ry 3d rhyme, C 3d consonant, V 3d vowel, SB 3d syllable boundary.



5.12 In this model, the serial order of phonemes during recall is governed by the cyclical activity of the syllable template. As in the CQ models discussed above, this template is formally a time-varying vector, and acts as a kind of control signal. However, the template does not lead to the activation of specific responses, as in Houghton (1990). Instead, its states are associated with whole classes of responses (phonemes). Which of the set of possible phonemes is to be produced is specified by a separate "content" input. This factoring of serial order information into a separate system means that the endogenous dynamical signals used in such CQ models as Houghton, 1990, Houghton et al., 1994, does not have to be repeatedly represented for every sequence learnt. This represents a considerable simplification, and is only possible in cases where the set of sequences to be learnt conforms to some underlying pattern which can be abstracted; in other words, where there is a grammar.

5.13 Whether such principles can be extended to sequencing outside the linguistic domain depends on whether other forms of action sequencing are susceptible to grammatical analysis, i.e., whether particular sequences can be seen as instantiations of an underlying schema defined, at least in part, in terms of variables. Error data from slips of action in normal subjects (Norman, 1981; Reason, 1984), and acquired disorders of action planning (e.g., Lhermitte, 1983; Luria, 1973; Schwartz, Reed, Montgomery, Palmer, & Mayer, 1991) support the idea that the kinds of sequencing principles we have

described can be applied to action sequences generally. For instance, transposition errors in routine actions are commonly found in patients with frontal apraxia. According to Schwartz et al. (1991), such patients are frequently recorded to put on their shoes before their socks, or to put toothpaste onto a tooth brush after having brushed their teeth. Cooper et al. (1994) describe a "hybrid" symbolic-connectionist model of routine action (based on Norman & Shallice, 1986) aimed at understanding such action slips. Routine actions are controlled by schemata activated in parallel, which compete for control of action on the basis of their activation values. Serial behaviour emerges from the model due to schemata being inhibited once their corresponding goals have been achieved. Elements of schemata contain variables which need to be given specific values during execution, for instance "arguments" representing the object on which an action is to be performed. Argument selection is based on activation levels of (representations) of objects, and how well they fit a "feature specification" associated with the schema. This is similar to the mechanism of phoneme selection by the syllable template (or "schema") in the Hartley and Houghton (in press) verbal recall model.

5.14 In addressing the issue of grammar last we do not intend to suggest that schema based sequencing is particularly exceptional or rare, though we do believe that the human capacity for it is greatly developed compared with other animals. Indeed, it is nongrammatical sequencing, as found in the single-trial learning of lists of items all belonging to the same class, that may be the comparatively unusual behaviour. The reason these models are addressed last is simply that they are the most complex, and the way in which they explain particular data presupposes that, independently of the operation of the schemata for order, groups of competing responses are being activated in parallel. Why should that be? The explanation we provide is that this parallelism represents one aspect of a more "primitive" form of sequencing. The development of schema based sequencing has not supplanted these basic mechanisms, rather it operates in conjunction with them. This has benefits. First of all, one should be aware that English is a somewhat unusual language in that its word order is highly constrained ("schematised"). Many other languages show much "freer" word order, by which it is meant that constituents in a sentence with a given meaning are not bound to appear in a single particular order (Givon, 1979). Of course, in any actually spoken utterance of such a language, all constituents do appear in some definite order. What determines this order, if it is not completely specified by grammatical schemata, and how does this ordering interact with grammatical ordering? One possibility is that all influences on order, grammatical or otherwise, act on the same competitive queueing output system. The resultant order reflects the relative strengths of these influences, with more strongly activated items appearing earlier (Prentice, 1966; MacWhinney, 1977; Sridhar, 1989). Another benefit of this interactive view of sequencing is that if control by grammatical schemata becomes disorganised or weakened (as may be the case for instance in agrammatism, Saffran, Schwartz, & Marin, 1980; or the frontal apraxia discussed above), the simpler, competition based, mechanisms still ensure that serial behaviour is possible, as long as concrete responses are activated. By contrast, production system models based on symbolic action grammars plus serial recursive processes (e.g. Houghton & Isard,

1987) are completely dependent on the grammar for the specification of the order of actions. If this mechanism breaks down, no behaviour can be produced.

6. Conclusions

"I have devoted so much time to ... the problem of syntax ... because the problems raised by the organization of language seem to me to be characteristic of almost all other cerebral activity...Not only speech, but all skilled acts seem to involve the same problems of serial ordering, even down to the temporal coordination [of] such a movement as reaching and grasping. Analysis of the nervous mechanisms underlying order in the more primitive acts may contribute ultimately to the solution even of the physiology of logic" (Lashley, 1951, p. 122).

6.1 We began this paper with Lashley's rejection of associative chaining as a basis for a neuropsychological theory of serial order, and his tentative suggestions for an alternative based on parallel response activation, and "schemata for action". Since Lashley wrote his article, a great deal has been learned about sequential behaviour, particularly in the linguistic domain. Various aspects of the behavioural data are briefly reviewed above, and we believe they support Lashley's view that models based on associative chaining are "doomed to failure". However alternative, neurally plausible, accounts of serial order compatible with a broad range of behavioural data have been few and scattered. The models described in this paper have attempted to develop a particular line of research compatible with Lashley's insights (e.g., Dell, 1986, 1988; Grossberg, 1978; Houghton, 1990; Mackay, 1987; Rumelhart & Norman, 1982; Shallice, 1972). The central interest of the current paper lies in its attempt to integrate these various strands into a coherent theory of serial order, applicable over a wide range of cases. In doing this we have built up a series of models, beginning with the basic problem of parallel response competition and its possible resolution into serial action. We then considered how a simple mechanism capable of resolving competition might be exploited to generate serial behaviour from memory rather than environmental stimulation. Models capable of this were discussed and their basic features adumbrated. It was claimed that such models provided unique insight into certain common error patterns. For instance, the models have a "default" mode of operation whereby actions are not repeated. Behaviour thus has a built in tendency to spontaneously "move on". This makes repetition, which might sometimes be necessary, a problem, and it was proposed that repetitive behaviour represents a specific mode, which must be engaged and disengaged. This mode might function by temporarily disabling the inhibitory feedback used in the normal model. This proposal has interesting empirical consequences, some of which have found support.

6.2 Following on, we identified certain limitations of these basic models and discussed ways in which they can be overcome, without abandoning the dynamical features of the models which make them so attractive. It was suggested that hierarchical models of serial order need not have dynamic properties only at the "terminal" or output level, but that control or chunk nodes at higher levels could change their pattern of activity in regular ways during learning and recall, in a manner reminiscent of the neural central pattern generators or endogenous "clocks" found in many species (Pearson, 1993; Treisman,

1994). This considerably increases the power of these models, and permits an expanded learning capacity. This has been exploited in the extension of the models to the domain of short-term memory. Finally, we came to what Lashley referred to as "schemata for action" and discussed how these schemata could be integrated with the models developed so far. This has permitted the application of these ideas to complex phenomena in language and other forms of action. Yet, even at this stage, the models retain the stamp of their particular origin, in that parallel response competition and the means of its resolution remain central explanatory mechanisms.

6.3 We conclude then that Lashley's original insights into serial order in human behaviour, largely based on everyday observations, remain valid. Perhaps most importantly, we believe that the work reviewed above provides concrete support for Lashley's conviction, expressed in the final quotation above, that similar ordering principles operate in many superficially different domains.

Acknowledgements

The authors are grateful to Steve Keele, Don MacKay, Rik Henson, and Steve Jackson for valuable comments on a previous version of this paper. We would also like to thank our colleagues Dave Glasspool, Tim Shallice, Steve Tipper, Neil Burgess and Gordon Brown for numerous discussions of matters raised in the paper.

Notes

<1> MacKay's models distinguish between "priming" and "activation". However, this (and other) complications will be left aside for simplicity's sake.

References

Aldridge, J.W., Berridge, K.C., Herman, M., & Zimmer, L., (1993). Neuronal coding of serial order: Syntax of grooming in the neostriatum. *Psychological Science*, 4, 391-395.

Allport, A., (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. In H. Heuer & A.F. Sanders, (Eds.). *Perspectives on perception and action*. Hillsdale, NJ: Erlbaum.

Amit, D.J., Sagi, D., & Usher, M., (1990). Architecture of attractor neural networks performing cognitive fast scanning. *Network*, 14, 189-216.

Ans, B., Coiton, Y., Gilhodes, J-C., & Velay, J-L., (1994). A neural network model for temporal sequence learning and motor programming. *Neural Networks*, 7(9), 1461-1476.

Baars, B.J., Motley, M.T., & Mackay, D.G., (1975). Output editing for lexical status in

artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 14, 382-391.

Baddeley, A.D., (1968). How does acoustic similarity influence short-term memory? *Quarterly Journal of Experimental Psychology*, 18, 362-365.

Baddeley, A.D., Gathercole, S., Bishop, D., & Papagno, C., (submitted). The phonological loop as a language learning device. *Psychological Review*.

Baddeley, A.D, Thomson, N., & Buchanan, M., (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14, 575-589.

Bairaktaris, D., (1992). A speech-based connectionist model of human short-term memory. *Proceedings of the 14th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Berridge, K.C., & Wishaw, I.Q., (1992). Cortex, striatum and cerebellum: Control of serial order in a grooming sequence. *Experimental Brain Research*, 90, 275-290.

Bisiacchi, P.S., Cipolotti, L., & Denes, G., (1989). Impairments in processing meaningless verbal material in several modalities: The relationship between short-term memory and phonological skills. *Quarterly Journal of Experimental Psychology*, 41A, 292-320.

Burgess, N., (1995). A solvable connectionist model of immediate recall of ordered lists. In G. Tesauro, D.S. Touretzky, & T.K. Leen (Eds.). *Advances in Neural Information Processing Systems*, 7, Cambridge, Mass.: MIT Press.

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

Chomsky, N., (1957). *Syntactic Structures*. The Hague: Mouton.

Church, R.M., & Broadbent, H., (1990). Alternative representations of time, number, and rate. *Cognition*, 37, 55-81.

Coles, M.G.H., Gratton, G., Bashore, T.R., Eriksen, C.W., & Donchin, E., (1985). A psychophysiological investigation of the continuous low model of human information processing. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 529-553.

Colombo, M., Eickhoff, A.E., & Gross, C.G., (1993). The effects of inferior temporal and dorsolateral frontal lesions on serial-order behavior and visual imagery in monkeys. *Cognitive Brain Research*, 1, 211-217.

Cooper, R., Shallice, T., & Farrington, J., (1994). Symbolic and continuous processes in

the automatic selection of actions. *Technical Report No. UCL-PSY-ADREM-TR11*, Dept. of Psychology, University College London.

Conrad, R., & Hull, A.J., 1964, Information, acoustic confusion and memory span. *British Journal of Psychology*, 55, 429-432.

Cowan, N., Day, L., Saults, J.S., Keller, T.A., Johnson, T. & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language*, 31, 1-17.

Cutler, A., (Ed). (1982). *Slips of the Tongue*. The Hague: Mouton.

Dehaene, S., Changeux, J-P., & Nadal, J-P., (1987). Neural networks that learn temporal sequences by selection. *Proceedings of the National Academy of Sciences, USA*, 84, 2727-2731.

Dell, G.S., (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, 93, 283-321.

Dell, G.S., (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, 25, 124-142.

Dell, G.S., & Reich, P., (1981). Stages in sentence production: an analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, 20, 611-629.

Ellis, A.W., (1979). Slips of the pen. *Visible Language*, 13, 265-282.

Ellis, A.W., (1980). Errors in speech and short term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behavior*, 19, 624-634.

Ellis, A.W., Young, A.W., & Flude, B.M., (1987). "Afferent dysgraphia" and the role of feedback in the motor control of handwriting. *Cognitive Neuropsychology*, 4, 465-486.

Ellis, N.C., (submitted). Sequencing in second language acquisition: Phonological memory, chunking and points of order. *Studies in Second Language Acquisition*.

Elman, J.L., (1990). Finding structure in time. *Cognitive Science*, 14, 179-211.

Eriksen, C.W., Coles, M.G.H., Morris, C.L.R., & O'Hara, W.P., (1985). An electromyographic examination of response competition. *Bulletin of the Psychonomic Society*, 23, 165-168.

Eriksen, B.A., & Eriksen, C.W., (1974). Effects of noise letters on the identification of a target letter in a nonsearch task. *Perception and Psychophysics*, 16, 143-149.

Estes, W.K., (1972). An associative basis for coding and organisation in memory. In A.W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, DC; Winston.

Fountain, S.B., Henne, D.R., & Hulse, S.H., (1984). Phrasing cues and hierarchical organization in serial pattern learning by rats. *Journal of Experimental Psychology: Animal Behaviour Processes*, *10*, 30-45.

Garrett, M.F., (1976). Syntactic processes in sentence production. In R.J. Wales & E. Walker (Eds.), *New Approaches to Language Mechanisms*. Amsterdam: North Holland.

Gathercole, S.E., & Baddeley, A., (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, *28*, 200-213.

Gathercole, S.E., & Baddeley, A., (1993). *Working Memory and Language*. Hove: Erlbaum.

Gathercole, S.E., Willis, C.S., Emslie, H., & Baddeley, A., (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, *12*, 349-367.

Givon, T., (1979) *On Understanding Grammar*. London: Academic Press.

Glasspool, D.W., (1995). Competitive queueing and the articulatory loop: An extended network model. In J. Levy, D. Bairaktaris, J. Bullinaria, & D. Cairns (Eds.), *Connectionist Models of Memory and Language*. London: UCL Press.

Glasspool, D.W., Houghton, G., & Shallice, T., (1995). Interactions between knowledge sources in a dual-route connectionist model of spelling. In L.S. Smith & P.J.B. Hancock (Eds.), *Neural Computation and Psychology*. London: Springer-Verlag.

Greenberg, J.H., (1978). Some generalizations concerning initial and final consonant clusters. In J.H. Greenberg (Ed.), *Universals of Human Language, vol. 2: Phonology*. Stanford, CA.: Stanford University Press.

Grossberg, S., (1978). Behavioral contrast in short term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology*, *17*, 199-219.

Hartley, T., & Houghton, G., (1995). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language*.

Henson, R., Norris, D , Page, M., & Baddeley, A.,(In Press). Unchained memory: error

patterns rule out chaining models of immediate, serial recall. *Quarterly Journal of Experimental Psychology*.

Hinde, R.A., (1970). *Animal behavior: A synthesis of ethology and comparative psychology*. New York: MacGraw-Hill.

Hitch, G., Burgess, N., Towse, J., & Culpin, V., (in press). Temporal grouping effects and working memory: the role of the phonological loop. *Quarterly Journal of Experimental Psychology*.

Hitch, G., Burgess, N., Shapiro, J., Culpin, V., & Malloch, M., (1995). Evidence for a timing signal in verbal short-term memory. Paper presented at the meeting of the Experimental Psychology Society, University of Birmingham, UK.

Houghton, G., (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish & M. Zock (Eds.), *Current research in natural language generation*. London: Academic Press.

Houghton, G., (1994). Inhibitory control of neurodynamics: Opponent mechanisms in sequencing and selective attention. In M. Oaksford & G.D.A. Brown (Eds.), *Neurodynamics and psychology*. London: Academic Press.

Houghton, G., Glasspool, D.W., & Shallice, T., (1994). Spelling and serial recall: Insights from a competitive queueing model. In G.D.A. Brown & N.C. Ellis (Eds.), *Handbook of spelling: Theory, process and intervention*. Wiley: Chichester.

Houghton, G., Hartley, T., & Glasspool, D.W., (In Press). The representation of words and nonwords in short-term memory: Serial order and syllable structure. To appear in S.E. Gathercole, (Ed.), *Models of Short-Term Memory*. Erlbaum.

Houghton, G., & Isard, S., (1987). Why to speak, what to say, and how to say it: Modelling language production in discourse. In P.E. Morris (Ed.), *Modelling Cognition*. Chichester: Wiley.

Houghton, G., & Tipper, S.P., (1994). A model of inhibitory mechanisms in selective attention. In D. Dagenbach & T. Carr (Eds.), *Inhibitory Mechanisms in Attention Memory and Language*. San Diego: Academic Press.

Houghton, G., & Tipper, S.P., (In Press). Inhibitory mechanisms of neural and cognitive control: Applications to selective attention and sequential action. *Brain and Cognition*.

Hulme, C., Maughan, S., & Brown, G.D.A., (1991). Memory for familiar and unfamiliar words: evidence for a long-term memory contribution to short-term span. *Journal of Memory and Language*, 30, 685-701.

Ingle, D., (1972). Selective choice between double prey objects by frogs. *Brain Behavior Evolution*, 7, 127-144.

Jackson, G.M., Jackson, S.R., Harrison, J., Henderson, L., & Kennard, C., (in press). Serial reaction time learning and Parkinson's disease: Evidence for a procedural learning deficit.

Jensen, A.R., & Rohwer, W.D., (1965). What is learned in serial learning? *Journal of verbal learning and verbal behavior*, 4, 62-72.

Jordan, M.I., (1986). Serial order: A parallel distributed approach. *ICI report 8604*, Institute for Cognitive Science, University of California, San Diego.

Keele, S.W., & Jennings, P.J., (1992). Attention in the representation of sequence: Experiment and theory. *Human Movement Studies*, 11, 125-138.

Kermadi, I., Jurquet, Y., Arzi, M., & Joseph, J.P., (1993). Neural activity in the caudate nucleus of monkeys during spatial sequencing. *Experimental Brain Research*, 94, 352-356.

Kesner, R.P., & Novak, J.M., (1982). Serial position curve in rats: Role of the dorsal hippocampus. *Science*, 218, 173-175.

Konishi, M., (1985). Birdsong: From behavior to neuron. *Annual Review of Neuroscience*, 8, 125-170.

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

Lhermitte, F., (1983). Utilisation behaviour and its relation to lesions of the frontal lobes. *Brain*, 106, 237-255.

Luria, A.R., (1973). *The Working Brain*. London: Penguin .

Mackay, D.G., (1970). Spoonerisms: the structure of errors in the serial order of speech. *Neuropsychologia*, 8, 323-350.

Mackay, D.G., (1972). The structure of words and syllables: evidence from errors in speech. *Cognitive Psychology*, 3, 210-227.

Mackay, D.G., (1987). *The organization of perception and action*. New York: Springer Verlag.

MacNeilage, P.F., (1964). Typing errors as clues to serial ordering mechanisms in language behaviour. *Language and Speech*, 7, 144-159.

- MacWhinney, B., (1977). Starting points. *Language*, 53, 152-168.
- Marler, P., (1991). The instinct to learn. In S. Carey & R. Gelman (Eds.) *The epigenesis of mind: Essays on biology and cognition*. NJ: Erlbaum. Reprinted in M.H. Johnson (1993) (Ed.), *Brain development and cognition: A reader*. Oxford: Blackwell.
- Neumann, O., (1987). Beyond capacity: A functional view of attention. In H. Heuer & A.F. Sanders, (Eds.), *Perspectives on perception and action*. Hillsdale, NJ: Erlbaum.
- Nissen, M.J., & Bullemer, P.T., (1987). Attentional requirements for learning: Evidence from performance measures. *Cognitive Psychology*, 19, 1-32.
- Norman, D., (1980). Categorization of action slips. *Psychological Review*, 88, 1-15.
- Norman, D., & Shallice, T., (1986). Attention to action: Willed and automatic control of behavior. In R. Davidson, G. Schwartz, & D. Shapiro, (Eds.), *Consciousness and Self Regulation*, Vol. 4, New York: Plenum.
- Pearson, K.G., (1993). Common principles of motor control in vertebrates and invertebrates. *Annual Review of Neuroscience*, 16, 265-297.
- Paulesu, E., Frith, C.D., & Frackowiak, R.S.J., (1993). The neural correlates of the verbal component of working memory. *Nature*, 362, 342-344.
- Posner, M.I., & Cohen, Y.A., (1984). Components of visual orienting. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and Performance X*. Hillsdale, N.J.: Erlbaum.
- Prentice, J.L., (1966). Response strength of single words as an influence in sentence behavior. *Journal of Verbal Learning and Verbal Behavior*, 5, 429-433.
- Reason, J.T., (1984). Lapses of attention. In W. Parasuraman, R. Davies, & J. Beatty, (Eds.), *Varieties of Attention*. Orlando: Academic Press.
- Rosenbaum, D.A., (1991). *Human Motor Control*. San Diego: Academic Press.
- Rumelhart, D.E., & McClelland, J.L., (1986). On learning the past tenses of English verbs. In J.L. McClelland & D.E. Rumelhart (Eds.), *Parallel Distributed Processing, Vol. 2: Psychological and Biological Models*. Cambridge, Mass.: MIT Press.
- Rumelhart, D.E., & Norman, D., (1982). Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science*, 6, 1-36.
- Saffran, E.M., Schwartz, M.F., & Marin, O.S.M., (1980). The word order problem in agrammatism II: Production. *Brain and Language*, 10, 263-280.

Schwartz, M.F., Reed, E.S., Montgomery, M., Palmer, C., & Mayer, N.H., (1991). The quantitative description of action disorganisation after brain damage: A case study. *Cognitive Neuropsychology*, 8, 381-414.

Seidenberg, M., & McClelland, J.L., (1989) , A distributed, developmental model of word recognition and naming. *Psychological Review*, 98, 523-568.

Selkirk, E., (1984). On the major class features and syllable theory. In M. Aronoff, & R.T. Oehrle, (Eds.), *Language sound structure: Studies in phonology presented to Morris Halle by his teacher and students*. Cambridge, Mass.: MIT Press.

Shallice, T., (1972). Dual functions of consciousness. *Psychological Review*, 79, 383-393.

Shallice, T., Glasspool, D., & Houghton, G., (in press). Can neuropsychological evidence inform connectionist modelling? Analyses from spelling. *Language and Cognitive Processes*.

Shattuck-Hufnagel, S., (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W.E. Cooper & E.C.T Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garret*. Hillsdale, NJ: Erlbaum.

Speidel, G.E., (1993). Phonological short-term memory and individual differences in learning to speak: A bilingual case study. *First Language*, 13, 69-91.

Sridhar, S.N., (1989). Cognitive structures in language production: A crosslinguistic study. In B. MacWhinney & E. Bates (Eds.), *The Crosslinguistic Study of Sentence Processing*. Cambridge: Cambridge University Press.

Stroop, J.R., (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.

Terrace, H.S., (1991). Chunking during serial learning by a pigeon: I. Basic evidence. *Journal of Experimental Psychology: Animal Behaviour Processes*, 17, 81-93.

Treiman, R., (in press). Errors in short-term memory for speech: A developmental study.

Treiman, R., & Danis, C., (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, 145-152.

Treisman, M., Cook, N., Naish, P.L.N., & McCrone, J.K., (1994). The internal clock: electroencephalographic evidence for oscillatory processes underlying time perception. *Quarterly Journal of Experimental Psychology*, 47A, 241-289.

Veneri, A., Cubelli, R., & Caffarra, P., (1994). Perseverative dysgraphia: A selective disorder in writing double letters. *Neuropsychologia*, 32, 923-931.

Wickelgren, W.A, (1969). Context sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, 76, 1-15.

Wilberg, R.B., (1990). The retention and free recall of multiple movements. *Human Movement Science*, 9, 437-479.