SERIOL Reading

Carol Whitney  
Department of Linguistics  
University of Maryland  
College Park, MD  

Piers Cornelissen  
Department of Psychology  
University of York  
UK
Abstract

In recent years, there has been a growing interest in the issue of orthographic processing in visual word recognition. In this article, we summarize recent work on the SERIOL model of letter-position encoding. We respond to criticisms of the model, present new data on letter priming, compare the ability of the SERIOL and other models to explain this and related data, and discuss the SERIOL model in the broader contexts of visual object recognition and dyslexia.
Following the publication of the SERIOL model of letter-position encoding (Whitney, 2001a), several alternative models have been presented, and an increasing number of experimental investigations have been performed in this area. The overall goals of this paper are to respond to some of these articles, and to summarize recent work on the SERIOL model. We address criticisms of the SERIOL model (Dehaene, Cohen, Sigman & Vinckier, 2005; Goswami & Ziegler, 2006), present new experimental data on letter priming, and compare SERIOL to other models in their ability to explain these and other experimental results. A commentary on one competing model, SOLAR (Davis, 1999; Davis & Bowers, 2006), has been presented elsewhere (Whitney, 2006a). Therefore, we concentrate on two other alternative models (Dehaene et al., 2005; Grainger & van Heuven, 2003).

We begin by defining the problem of letter-position encoding, and discuss why this problem is of practical and theoretical interest to the broader domains of dyslexia and visual object recognition. Next, we briefly review the three models under consideration. We then compare the models in various contexts, as follows. First, we review letter-priming experiments, present new data, and discuss whether the models can account for these data. Second, we discuss the models’ ability to explain letter perceptibility patterns. Third, we discuss serial processing. Fourth and fifth, we consider the robustness of the proposed representations for orthographic lexical access and phonological processing, respectively. We conclude with a brief discussion of the implications for object recognition and dyslexia.
Letter-Position Encoding

When a visual image first arrives at the cortex, it is in the form of a retinotopic encoding. To process the image of a word, this spatial representation must be transformed to an encoding of the constituent letters’ identities and positions. There is a growing consensus that reading involves multiple processing routes. On the lexical route, a letter string directly accesses lexical representations. On the sub-lexical route, a string is mapped onto a phonological representation, which then contacts lexical representations. Thus one aspect of the problem of letter-position encoding is to define how a retinotopic representation is converted into abstract encoding of letter order, which provides input to the lexical and sub-lexical routes. A further aspect is to define how this encoding then activates lexical representations along the lexical route. Note that the problem of letter recognition is not addressed. Rather, the question is more abstract – given the ability to identify letters, how are the relationships between letters derived from a spatial encoding, and represented to allow lexical and phonological processing?

This question is of practical interest for the problem of developmental dyslexia. Although it has been widely assumed that dyslexia stems from a core phonological deficit, this is not necessarily the case (Castles & Coltheart, 2004). There is evidence that aspects of visual processing are abnormal in many dyslexics (e.g., Buchholz & Aimola Davies, 2006; Cornelissen, Richardson, Mason & Stein, 1995; Facoetti & Molteni, 2001; Livingstone, Rosen, Drslane & Galaburda, 1991, Sperling, Zhong-lin, Manis, & Seidenberg, 2003). If the visual system cannot provide an accurate encoding of letter order, this will interfere with the ability to learn the relationships between letters and
sounds, leading to an impaired ability to read pseudowords and to perform phonological-awareness tasks (Whitney & Cornelissen, 2005). Thus a visual deficit could cause difficulties in phonological processing, in the absence of a core phonological deficit. Furthermore, dyslexics generally do not display phonological problems in languages having highly consistent mappings between graphemes and phonemes (Miles, 2000). Rather, their reading is slow but accurate, suggesting that they have not learned to visually process letter strings in an efficient manner. In order to understand what could go awry in dyslexics’ visual processing, it is first necessary to understand normal visual processing in reading.

The question of letter-position encoding is also of interest to the field of visual object recognition. An abstract orthographic encoding requires a representation that is not tied to retinal location, but which encodes spatial relationships between the letters. These problems of location invariance and representing sub-part relationships are also highly relevant to visual object recognition. Understanding how the brain solves these problems in the restricted domain of visual word recognition may shed light on basic mechanisms of visual processing.

**Review of Models**

[Figure 1 about here.]

The three models under consideration are summarized in Figure 1. They will each be presented in a bottom-up manner. For brevity, we will simply specify the proposed
representations and not address the motivations for choosing them. For the SERIOL model, these issues are addressed in detail elsewhere (Whitney & Berndt, 1998; Whitney, 2001a). For clarity, capitalized boldface will be used to denote a model unit encoding an item, while italics will be used for a stimulus. For example, **CART** denotes a word unit that recognizes the string *cart*.

**SERIOL**

The SERIOL model is a theoretical framework that specifies how information is represented at each of its processing layers, and how the representation in each layer is transformed into the representation at the next layer. The model focuses on bottom-up processing, but this is not meant to rule out top-down interactions.

The SERIOL model is comprised of five layers: edges, features, letters, open-bigrams, and words. The first two layers are retinotopic, while the latter three are abstract. For the retinotopic layers, the term *activation level* will be used to denote the total amount of neural activity devoted to representing a letter within a given layer. A letter’s activation level increases with the number of neurons representing that letter and their firing rate. For the abstract layers, the term *activation* will denote the activity level of a representational unit in a given layer.

The edge layer models V1/V2. In these areas, the rate of spatial sampling (acuity) is known to sharply decrease with increasing eccentricity. This is modeled by the
assumption that activation level decreases as distance from fixation increases. This pattern is termed the *acuity gradient*.

The feature layer models V4. Based on learned, hemisphere-specific processing, the acuity gradient of the edge layer is converted to a monotonically decreasing activation gradient (dubbed the *locational gradient*) in the feature layer. That is, activation level is highest for the first letter, and decreases across the string. Hemisphere-specific processing is necessary because, for a fixated word, the acuity gradient does not match the locational gradient in the first half of the word (i.e., acuity *increases* from the first letter to the fixated letter, whereas the locational gradient *decreases* across the string), whereas the acuity gradient and locational gradient match in the second half of the word (i.e., both decreasing). Strong directional lateral inhibition is required in the hemisphere contralateral to the first half of the word, in order to invert the acuity gradient. This processing is summarized in Figure 2.

At the letter layer, corresponding to posterior fusiform gyrus, the locational gradient interacts with oscillatory letter units, creating sequential firing. That is, the letter unit encoding the first letter fires, then the unit encoding the second letter fires, etc., on the time scale of about 10 ms per letter. This interaction also causes letter units to have varying activations, where activation generally decreases across the string, but rises for the final letter. Note that the creation of a serial encoding is the key point of abstraction. The retinotopic representation is mapped onto a temporal one, providing a location-
invariant representation of letter order. This serial encoding provides input to both the lexical and sub-lexical routes. The processing in the following layers is specific to the lexical route.

At the open-bigram (Grainger & Whitney, 2004) layer, corresponding to left middle fusiform, units recognize pairs of letter units that fire in a particular order. That is, open-bigram unit $XY$ is activated when letter-unit $X$ fires before $Y$, where the letters $x$ and $y$ were not necessarily contiguous in the input string. The activation of an open-bigram unit decreases with increasing time between the firing of the constituent letter units. The open-bigram units connect to the word layer, via weighted connections. The input to a word unit is the dot-product of the bigram vector and its weight vector.

Whitney (2004a) specified several refinements to the open-bigram representation first presented in Whitney and Berndt (1998). (1) Originally, open-bigram activations also depended on the constituent letter units’ activations (which depend on string position). This assumption has been dropped. Experimental data that were originally explained by positional variations in bigram activations are now explained directly by seriality (Whitney, 2004a; Whitney, 2006b). (2) The representation is now taken to include edge units, which explicitly encode the first and last letters. For example, the encoding of the stimulus CART would be $*C$, $CA$, $AR$, $CR$, $RT$, $AT$, $CT$, and $T*$, where $*$ represents an edge. (3) Normalization of weights is assumed, such that weights to shorter words are larger than weights to longer words. For example, the weights on the connections from $CA$, $AN$, and $CN$ to $CAN$ are larger than to $CANON$. Hence, the string can would
activate **CAN** more than **CANON**. Changes (2) and (3) replace the original assumption that the bigram encoding the first and last letter possessed special properties.

**Grainger and van Heuven (2003)**

In this model, an alphabetic array is converted in parallel into an open-bigram encoding, where open-bigram activations are either 1 or 0. Open bigrams are activated by letter pairs having up to two intervening letters. The open-bigram units then connect to word units.

The alphabetic array is a retinotopic encoding, in which there are separate representations of a given letter at different retinal locations. The way in which the array is converted into the open-bigram encoding is not specified. However, it can be reasoned that such recognition would require retinotopic open-bigram detectors. For example, an **XY_0** detector might respond to an **X** at fixation, and a **Y** at locations to the right of fixation. An **XY_1** detector might respond to an **X** at 1 degree to the right of fixation, and a **Y** at locations to the right of this. Note that multiple location-specific open-bigram detectors are necessary to assure that **X** occurred to the left of **Y**. The retinotopic open-bigrams would then connect to abstract open bigrams. For example, all **XY_n** would connect to an abstract **XY** detector. Thus the model can be summarized as consisting of four layers: retinotopic letters, retinotopic open-bigrams, open-bigrams, and words.

**Dehaene et al. (2005)**

This model also starts out with retinotopic letters, and retinotopic open-bigrams. However, the retinotopic open-bigrams then activate four-grams, which connect to word
units. The four-grams would also have to be retinotopic (in order to recognize open-bigrams in a particular order), but then the model would not include a location-invariant sub-lexical encoding of letter order. For example, under the given scheme, learning a novel four-letter word presented at fixation would not allow recognition of that word some distance from fixation. The authors have previously claimed that the left fusiform gyrus houses a location-invariant encoding of letter order (McCandliss, Dehaene & Cohen, 2003), so we will assume that the retinotopic four-grams then activate abstract four-grams.

Summary

The models differ in three key ways as illustrated in Fig. 1. (1) Abstraction occurs at different levels of processing. For the SERIOL, Grainger and van Heuven, and Deheane et al. models, the location-invariant encoding first occurs at the letter, open-bigram, and four-gram levels, respectively. Thus the SERIOL model is the only one that includes an abstract representation of individual letters. Such a letter unit can represent that letter in any retinal location, wherein timing firing binds positional information to the letter identity. In contrast, there are separate letter units for a given letter at different retinal locations in the other models.

(2) The SERIOL model includes a serial encoding, while the other models are purely parallel. The other models reflect the standard assumption that location invariance is achieved directly via an increase in receptive-field size and complexity. In contrast, space is mapped onto time in the SERIOL model to create an abstract, invariant representation.
This explicit abstraction mechanism allows an invariant representation to be achieved at a lower level of processing (i.e., individual letters) than the other models.

(3) In the SERIOL model, but not the others, activation patterns play an important role. A monotonic activation gradient at the feature level is necessary to create the serial encoding. Due to the non-monotonic acuity gradient at the edge level, hemisphere-specific processing is necessary to create this activation gradient. The serial letter encoding creates varying activation levels, and open-bigram activations encode the distance between the constituent letters. In the parallel models, graded activations are not specified; units are taken to either be on or off.

In comparing the models, we will first look at experimental results on non-pronounceable letter strings. Such stimuli isolate processing occurring at an orthographic, pre- lexical, non-phonological level. Priming experiments indicate the existence of abstract letter representations, while perceptual studies show patterns specific to orthographic encoding.

**Letter Priming**

Peressotti and Grainger (1995) performed a series of alphabetic-decision experiments that are very informative as to the nature of orthographic encoding at the letter level. The task was to identify whether a three-character stimulus consists entirely of letters. Targets were comprised of three consonants, while foils consisted of two consonants and a non-letter character, such as “&”. A briefly presented prime (33, 50, or 67 ms) and a mask preceded the stimulus. The goal was to investigate the effect of differing relationships between prime and target. To indicate such relationships, the prime’s letters will be
denoted by their position in the target, with “d” indicating a letter not in the target. For example, for the target GDK, the prime GDK is denoted 123, KGD is 312, and LDK is d23.

[Figure 3 about here.]
The results of the experiments are summarized in Figure 3. Several things are evident. For same-position primes (123), facilitation is robust and exposure duration has little effect. The other prime types give a reduced effect. For cross-position primes (312), facilitation is sensitive to exposure duration. Replacement primes (12d, 1d3, d23) induced a positional effect (12d produced significantly more facilitation than 1d3), while transposition primes (132, 321, 213) did not.

The fact that facilitation occurred for cross-position primes indicates that there are letter representations that can encode a given letter at different retinal locations and string positions. Indeed, based on prime duration effects, the authors concluded that there are two types of letter detectors, position-specific and position-independent, Position-specific detectors activated first, followed by the position-independent detectors. Whitney (2001b) argued that the position-specific effect may not actually stem from string position, but rather may be retinotopic. Although primes and targets were presented in different font sizes (so letters in the same string position in the prime and target appeared at different retinal locations), there may still have been retinotopic overlap between letters at the same position, due to feature detectors with receptive fields large enough to span the differing retinal locations.

[Figure 4 about here.]
Indeed the data are well approximated under the following assumptions. The prime activates a retinotopic encoding in parallel (corresponding to the feature layer of the SERIOL model), which decays over time. The retinotopic encoding activates abstract letter units serially at the rate of 10-15 ms a letter. For a 33 ms prime duration, only the first letter of the prime activates an abstract letter unit. For 50 ms, the first two letters of the prime activate abstract letter units, while all three abstract letter units are activated at 67 ms. The following values provide a reasonable fit to the data, shown in Figure 4. For prime durations of 33 or 50 ms, a retinotopic match (match on letter identity and position) provides 6 ms of facilitation per letter. For 67 ms, retinotopic priming is reduced to 3 ms per letter, reflecting decay of this early level of representation. An abstract match (match on letter identity) provides an additional 5 ms of facilitation per letter.

For cross-position primes, the seriality explains the monotonic increase with exposure duration. The seriality also explains the difference in positional effect between transposition and replacement primes, for which presentation duration was 50 ms. At this duration, only the first two positions of the prime activate abstract letter units. For transposition primes, the first two letters are present in the target in every condition (132, 321, or 213). Hence, all transposition primes provide one retinotopic and two abstract matches, giving no effect of the matched position. However, the situation is different for replacement primes. When the first two letters of the prime are in the target (12d), they both generate abstract matches. When the prime’s first two letters include one that is not in the target (1d3 or d23), an abstract match can only occur for the letter that is in the
target. Thus 12d gives two abstract matches, while 1d3 and d23 each yield only one abstract match, accounting for the observed positional effect. This leads to the prediction that the positional effect should disappear for a prime duration of 67 ms, because an abstract unit representing the prime’s third letter should also become activated. That is, all replacement primes should yield two abstract matches at this longer duration.

Thus these data are consistent with SERIOL’s proposal of sequential activation of abstract letter units. The other models cannot account for these data because they do not include abstract, location-independent letter units. Note that the phenomena cannot be explained by simply proposing that there are letter units with receptive fields spanning multiple string positions. Such units would lose important locational/positional information. There would have to be an additional mechanism for binding this information to such units, which neither of the parallel models has specified.

**Experiment**

This account predicts that it should be possible to observe priming across large differences in retinal location, because a given letter should activate the same abstract letter unit, independently of retinal location. To test this prediction, we performed an alphabetic-decision experiment in which the prime was presented bilaterally, and the stimulus was at fixation. Primes were of the form 123, 312, or ddd. Because same-position primes should no longer enjoy a retinotopic advantage, we expected no difference between the 123 and 312 conditions. Following the psychophysical approach of the original study, we used a small number of subjects and a large number of repetitions.
In order to create a relatively large distance between the prime and stimulus while allowing the stimulus to be displayed for more than 150 ms, we were forced to present the prime parafoveally. However, parafoveal presentation is problematic in that it likely produces less activation and less robust priming than a central prime. Indeed, in the initial analysis of the data, we found no effect of prime type. We therefore performed a further analysis to look for priming effects. Reaction times in two-alternative forced-choice experiments are known to be influenced by expectancy when the inter-trial interval is greater than 500 ms (Soetens, Boer & Hueting, 1985), as was the case in this experiment. Soetens et al. found that target reaction times are faster if the previous stimulus is a foil rather than a target, reflecting expectancy for alternation. Therefore, we evaluated possible interactions between expectancy and priming in our analyses.

Method

Subjects. Four adults, including one of the authors (C.W.), volunteered as subjects. All had normal or corrected-to-normal vision.

Materials. The target stimuli were 30 trigrams consisting of three different letters, formed by random combination of the following 17 consonants: B, C, D, F, G, H, K, L, M, N, P, R, S, T, V, X, and Z. None of the trigrams formed acronyms familiar to the subjects. Targets were preceded by one of three types of prime: same-position (123), cross-position (312), or absent (ddd).
There were 30 foil trigrams consisting of two consonants and one of the following 10 characters: @, <, >, ), (, %, *, ~, ?, and =. Each non-letter character occurred once in each string position. Foils were preceded by a three-consonant prime.

The 30 trigrams were divided into three groups, which were rotated through the three prime conditions in three blocks. Thus, an experimental session consisted of three blocks, each containing the 30 targets and 30 foils; across blocks, each target was preceded by a different prime type. Stimulus order was randomized in each block, and block order was randomized across sessions. Each subject completed four sessions on different days. Thus each subject saw 120 target trials for each of three prime conditions.

**Procedure.** The experiment was controlled by Presentation software, using a monitor with a 60-Hz refresh rate. Stimuli were presented in black on a white background. Primes, targets and foils were presented as capital letters in 16-point New Courier. Each trigram spanned 1 cm, subtending approximately 1° at a viewing distance of about 50 cm. A prime was centered at an eccentricity of 2°, and presented simultaneously in both visual fields. Subjects completed one practice session of three blocks before undertaking the four experimental sessions.

A trial started with a + in the center of the screen for 500 ms. Then the prime was displayed in both visual fields for 50 ms. Next a mask comprised of 25 hash marks was displayed for 17 ms. Then a foil or target was presented, which remained on the screen until there was a response or time out. Between trials, the screen was blank. Trial timing
was controlled by the computer, with a randomly varying inter-trial interval of 1000 to 1500 ms. After a block of 60 trials, the subject initiated the next block at will.

In a go/no-go procedure, the subject was instructed to fixate on the +, and to press the space bar if the subsequent trigram at fixation consisted entirely of letters. Pressing the space bar terminated the trial. Otherwise, the trial timed out after 750 ms.

[Insert Tables 1 and Figure 5 about here.]

**Results**

*Reaction Times.* Because a go/no-go procedure was used, reaction times were only available for targets. All data were analyzed. We first performed a one-way repeated-measures ANOVA on prime type, which was not significant, $F(2,6) < 1$.

We then investigated whether there was an interaction between prime type and expectancy. After discarding the first trial of each block, we divided the target trials by the type of the previous trial. As expected, this divided the data approximately in half for each subject in each prime condition. These results are given in Table 1 and Figure 5.

A two-way repeated-measures ANOVA (prime x previous-trial) revealed a significant interaction between prime and previous-trial, $F(2,6) = 3.5, p = .05$, while neither main effect was significant, prime: $F(2,6) < 2$; previous-trial: $F(1,3) < 1$.

[Insert Table2 about here.]
Examination of figure 5 shows that this interaction arose because there was no effect of prime type when the previous trial was foil, $F(2,6) < 1$, while there was an effect when the previous trial was target, $F(2,6) = 6.02$, $p < .05$. Planned pair-wise comparisons showed that this effect of prime type arose from a significant difference between the absent and same-position conditions, $F(1,3) = 10.94$, $p < .05$, and a marginally significant difference between the absent and cross-position conditions, $F(1,3) = 7.44$, $p = .07$, while there was no difference between the same- and cross-position conditions, $F < 2$. These results reflect the consistency of priming patterns across individual subjects, as shown in Table 2.

Accuracy. Accuracy was high. Misses occurred on 0.2% of the target trials, while false alarms occurred on 0.9% of the foil trials.

Discussion

We observed an interaction between prime type and the previous trial. When the previous trial was a foil (target expected), prime type had no effect. When the previous trial was a target (foil expected), same-position and cross-position primes were similarly facilitative (with respect to the absent condition) and did not differ from each other, in line with our expectations. This isolates the facilitation to the level of individual letters, as a cross-position prime (213) would not activate the same higher-order units (open bigrams, four-grams) as the target. Figure 5 shows that this facilitation compensated for the inhibitory cost in the absent condition of not expecting a target (Soetens et al., 1985).
The observed interaction between prime type and previous trial suggests that the expectation of a target overrode any priming effect. Although facilitation was less robust than the original study (Peressotti & Grainger, 1995), this was not wholly unexpected, as a parafoveal prime likely produces less activation than a central prime. Nevertheless, we observed significant priming effects. In the parallel models, letter representations are retinotopic, ruling out any effects of prime type at the letter level across large differences in retinal location. We demonstrated the existence of such effects, contrary to the parallel accounts, and consistent with the abstract letter representations in the SERIOL model.

**Perceptual Patterns**

Perceptual tasks provide another method for investigating orthographic encoding. In particular, it is of interest to note how perceptibility varies with string position, and how the pattern across string positions interacts with the alphanumeric status of the characters, visual field, and reading direction.

Two studies have examined perceptual patterns for letters versus non-alphanumeric characters in arrays of centrally presented stimuli (Hammond and Green, 1982; Mason, 1982), using a between-subjects design for the different stimulus types. Both studies found an external-character advantage for letters. That is, the first and last letters were processed more efficiently than the internal letters. Mason (1982) also showed an external-character advantage for number strings. However, both studies found that the advantage was absent for non-alphanumeric characters; the first and last characters were
processed the least well, in line with their lower acuity. Thus symbols that do not normally occur in strings show a different perceptual pattern than alphanumeric characters. Therefore, the external-letter advantage cannot be due to a general reduction of lateral inhibition, because such a low-level effect should be present for any type of character. Any account of letter-string processing should then explain how the external-letter advantage arises.

For three-letter strings briefly presented to a single visual field for left-to-right readers, the perceptual pattern depends on visual field. In the LVF, the first letter is perceived much better than the third, while the first and third letters are perceived equally well in the RVF (Hellige, Cohen & Eng, 1995; Legge, Mansfield, & Chung, 2001). This asymmetry reverses with reading direction, as shown by studies in Hebrew (Eviatar, 1995) and Urdu (Adamson & Hellige, 2006). Thus these patterns are clearly contingent on the demands of encoding letter strings for reading, and therefore any model of orthographic processing should account for them.

As described in more detail elsewhere (Whitney, 2001a; Whitney, 2006b; Whitney & Cornelissen, 2005), the SERIOL model explains these patterns. The external-letter advantage arises as follows. An advantage for the initial letter comes from the directional inhibition at the feature level, because the initial letter is the only letter that does not receive lateral inhibition. An advantage for the final letter arises at the letter level, because the firing of the last letter is not terminated by a subsequent letter. This processing is specific to letter strings, explaining the lack of external-character advantage
for non-alphanumeric characters. However, if the initial-letter advantage arises at the feature level (i.e., prior to character identification), how could it be specific to letters? Recall that the subjects seeing non-alphanumeric characters only saw that type of stimulus, under a between-subjects design (Hammond & Green, 1982; Mason, 1982). Thus string-specific mechanisms would not have been activated in such subjects. This suggests that an initial-character advantage for non-alphanumeric characters should emerge under a within-subjects design.

The LVF pattern of a strong advantage for the first over the third letter is explained by the proposed feature-level, left-to-right inhibition specific to the LVF/RH in left-to-right languages. In the absence of such inhibition, there is a shallower decrease in activation across the letters, yielding the more even RVF/LH pattern. For right-to-left languages, this directional inhibition should occur for the RVF/LH, explaining the reversal of trigram patterns with reading direction.

Whitney (2004b) proposed that visual-field asymmetries in lexical decision (Lavidor & Ellis, 2002; Young & Ellis, 1985) stem from these differing orthographic activation patterns. This theoretical account was later experimentally confirmed by using positional contrast manipulations to change activation patterns, resulting in reversal of the normal asymmetries (Whitney & Lavidor, 2004; 2005).

In the parallel models, activation levels (if specified) would only vary with acuity and number of neighboring letters (which would determine the amount of general, low-level lateral inhibition). Thus they cannot explain the differing patterns for alphabetic versus
non-alphanumeric characters, nor can they explain the hemifield patterns. Moreover, the SERIOL account led to precise, accurate predictions on how to reverse hemifield asymmetries in lexical decision (Whitney, 2004b; Whitney & Lavidor, 2004; 2005); the other models cannot explain these experimental results. The SOLAR model (Davis, 1999; Davis & Bowers, 2006) cannot explain these patterns or results either (Whitney, 2006a).

Next we contrast the models at higher levels of processing. We examine arguments for and against serial lexical access, and discuss the suitability of the proposed representations for lexical activation along the both the lexical and sub-lexical routes.

**Seriality**

Because lexical-decision experiments have indicated that reaction times do not vary with string length (Frederiksen & Kroll, 1976; Richardson, 1976), it has generally been assumed that lexical access occurs in parallel. However, this reasoning is problematic. It depends on the assumption that the amount of time that it takes to reach response criterion after all the letters have fired (the settling time) is independent of the number of letters. However, if the greater amount of bottom-up input from a longer word were to yield a decreased settling time, this could cancel or override the increased time it takes for all the letters to fire. Thus there could be no length effect, or even a reversed length effect, under a serial encoding.

Indeed, there are a number of studies consistent the idea that longer strings reduce settling time. In MEG studies, both Cornelissen *et al.* (2003) (4-, 6- and 8-letter words) and Tarkiainen *et al.* (1999) (single letters, 2-letter consonants and 4-letter words)
showed a systematic *reduction* with increasing string length in the latency of letter-string specific responses around ~150ms after stimulus presentation. Furthermore, a recent analysis (New, Ferrand, Pallier & Brysbaert, 2006) of lexical decision reaction times on a database of over 40,000 words (Balota *et al*., 2002) showed that RTs actually *decrease* as length increases from three to five letters; RTs are constant for words of five to seven letters, and RTs increase with length for seven or more letters. Hence, length has a facilitative, neutral, or inhibitory effect, depending on the range considered. It is highly unlikely that this variation reflects different methods of lexical access for words of different lengths. Thus the presence or absence of a length effect cannot be used to divine whether lexical access is serial or parallel.

Rather, in order to test whether strings are processed serially, temporality should be manipulated directly. Two such studies have provided evidence for seriality. Harcum and Nice (1975) presented a six-letter pseudoword for 70 ms, which was replaced by another six-letter pseudoword for 30 ms. Subjects were to report the letters perceived. Nine out of ten subjects reported the beginning letters of the first string and the final letters of the second string, with the crossover position between strings varying with subject. These results show that there was only enough time to process some of the letters of the first string before it changed to the second string, thereby providing direct evidence for serial processing.

Adelman & Brown (2006) presented a four-letter word for variable durations (0 to 42 ms, in increments of 6 ms), followed by a mask and a two-alternative forced choice between the target and a foil word. The foil was of various types: (1) single-replacement: d234,
1d34, 13d4, or 123d; (2) double-replacement: dd34, 1dd4, 12dd, or d23d; (3) transposition: 2134, 1324, 1243, or 4231. Thus the position of difference between the target and foil was parametrically varied in each of the foil types. Based on the interaction between accuracy, exposure duration and position of difference, they found that latency of letter information increased with string position. Information about letter identity and order was available earlier for the initial letter than the internal letters, and earlier for the internal letters than the final letter. These results are consistent with a read-out of letter information from left to right on a millisecond time scale. It is interesting to note that there was no final-letter advantage. This is consistent with the SERIOL account of the final-letter advantage, which depends on the ability of the final letter unit to fire for a longer period of time than the internal letters. At very brief presentation durations, the final letter is not reached or does not receive enough input to sustain prolonged firing, yielding no final-letter advantage.

Thus, there is no data in conflict with seriality, and direct evidence for serial processing. These results are predicted by the SERIOL model, and cannot be explained by purely parallel models. Furthermore, simulations presented in Whitney (2006b) show how serial activation of open-bigrams accounts for error patterns in aphasia (Berndt, Haendiges & Mitchum, 2005; Whitney & Berndt, 1998), and for both the presence and absence of positional effects in masked form priming under different conditions (Schoonbaert, & Grainger, 2004; Grainger, Granier, Farioli, van Assche & van Heuven, 2006).
Adequacy for Lexical Access

Dehaene et al. (2005) argued that an open-bigram encoding could not provide a robust enough encoding of letter order to support lexical recognition. They considered an on/off encoding with no edge units, arguing that such a representation fails to give a unique coding to strings with repeated letters. For example, the strings sense, ensse and sensse would all generate the same representation under these assumptions.

However, this argument does not apply to the open-bigram representation specified in Whitney (2004a), in which graded activations represent the separation of the component letters, edge units specify the external letters, and weights encode length information. To test the adequacy of this encoding, the string sensse was added to the word database of an implementation of the bigram and word levels of the model (Whitney, 2006b). With no change to the parameter settings, the stimulus sense yielded SENSE as the winning word unit, while sensse activated SENSSE as the winner. Thus an open-bigram encoding can differentiate between very similar strings with repeated letters, contrary to the claims of Dehaene et al.

Their solution to the ambiguity problem was to assume units encoding higher order sets of letters - bigrams activate four-grams, which activate words. However, such a representation means that there is zero similarity between similar words such as care and core. That is, the stimulus core would not activate the word-unit CARE at all, because it
would not activate the four-gram **CARE**. Intuitively, it is obvious that *core* is more similar to *care* than to a word matching no letters, yet their proposed encoding does not capture this similarity. While it could be argued that *core* partially activates the four-gram **CARE**, such partial activations were not specified in their model. Once partial activations are considered, open-bigrams can attain a finer encoding of letter order (than under an on/off constraint), and there is no reason to include a higher-order encoding. Thus their model is not viable as presented, while the open-bigram encoding presented in Whitney (2004a) is sufficiently robust to avoid ambiguity problems.

However, the open-bigram encoding proposed by Grainger and van Heuven (2003) *is* subject to such problems, because it uses on/off units. For example, experiments performed by Davis and Bowers (2006) indicate that 12d4 is more similar to 1234 than is 1d24. On/off open-bigrams cannot explain this pattern (because 12d4 and 1d24 both generate the same encoding), while graded open-bigrams do account for such a difference. As discussed in Whitney (2006a), Davis and Bowers (2006) incorrectly claimed that SERIOL model is inconsistent with their data. Their incomplete analysis was based on the outdated specification of bigram activations given in Whitney & Berndt (1998), in which activation varied with string position. Under the current parameterization (Whitney, 2004a), the SERIOL model does indeed predict their results for all conditions.
**Phonology**

Goswami and Ziegler (2006) argued that the demands of phonological processing should also be considered in constructing models of visual word recognition. They pointed out that parallel, open-bigrams do provide a suitable basis for learning grapheme-phoneme mappings, as open-bigrams are not phonologically relevant units.

However, their argument does not apply to the SERIOL model, because there is an abstract, serial, location-invariant representation of individual letters, which provides input to both the lexical and sub-lexical routes. The open-bigram representation is taken to be specific to the lexical route (Whitney & Cornelissen, 2005). It is assumed that the sub-lexical route divides the sequence of letters into graphemes, which are mapped to phonemes. (However, the details of this processing are beyond the scope of the model.) Indeed, there is mounting evidence that the phonological representation is computed serially (Carreras, Ferrand, Grainger & Perea, 2005; Roberts, Rastle, Coltheart & Besner, 2003), which meshes nicely with the proposed serial letter representation.

Despite the suitability of SERIOL’s representations for the demands of phonological processing, Goswami and Ziegler (2006) argued that “This solution ignores data showing that phonology affects the lexical route, such as body-neighborhood effects in lexical decision (Ziegler & Perry, 1998)”. It is unclear what is meant by this statement. First, Whitney (2004b) specifically discussed the data presented by Ziegler and Perry (1998), explaining in detail how the SERIOL model explains their findings. Second, the general
issue of interaction between the lexical and sub-lexical routes is orthogonal to the question of how letter order is encoded. Presumably, the lexical and sub-lexical routes converge onto the same lexical representations. The usual assumption of top-down connections from the lexical representations back to letters and phonemes would cause interaction between the routes, because orthographically-driven lexical information could then affect the phonemic representation, and vice versa. This connectivity pattern is independent of the issue of letter-position encoding.

In contrast to the SERIOL model, the parallel models do not include an abstract letter representation. Hence, they do not contain a level of representation suitable for phonological learning.

**Discussion**

We have seen that the SERIOL model, unlike the parallel models, explains priming and perceptual patterns at the letter level, is consistent with evidence for serial processing, and provides a representation suitable for phonological processing.

We indicated above that understanding letter-position encoding could potentially be useful for understanding dyslexia and visual object recognition. If the SERIOL model does provide a fairly accurate account of letter-position encoding, what would be the implications for these domains? We conclude by briefly discussing this question, starting with object recognition.
Representation of a letter string is the canonical example of analysis by parts, where subpart identities, not their visual details, determine the identity of an object (Farah, 1990). We will consider the implications of the SERIOL model for analytic processing in particular. If the proposed serial/oscillatory mechanism is indeed used in visual word recognition, it would have to stem from an innate mechanism for visual processing. Priming studies have indicated that attention is necessary for constructing an analytic, parts-based representation; in the absence of attention, only a more holistic representation is formed (Thoma & Davidoff, 2006; Thoma, Hummel & Davidoff, 2004). However, strings differ from objects in that strings are directional (\textit{dog} ≠ \textit{god}), while objects are not. Therefore, we propose that construction of an analytic, parts-based representation of an attended object relies on a serial read-out from the center of attention outwards, in all radial directions. Attention provides an activation gradient, even if the object is not a fixation. This attention gradient functions like the locational gradient to induce serial firing along each radial direction.

This proposal is consistent with single-cell studies of macaque V4. One study demonstrated the existence of cells with radial receptive-fields, as would be required for a radial serial read-out (Pigarev, Nothdruft & Kastner, 2002). The firing latency of such cells seems to increase with the eccentricity of the stimulus (Pigarev, pers. comm.), as would be required in mapping space to time. Another study (Connor, Preddie, Gallant & van Essen, 1997) showed that cells’ firing rates decrease as the distance to the locus of attention increases (providing evidence for an attention gradient), and that some cells are
very sensitive to the direction of the attentional locus (consistent with a radial representation originating at the center of attention).

Thus we propose that a parts-based representation is normally constructed from the center outwards. However, such a representation would not be suitable for encoding letter order. Hence we suggest that a key part of learning to read is to co-opt the normal radial mechanism to provide a serial encoding from the first to the last letter. This would require the creation of a suitable attention gradient. When fixating near the center of a string, attention would have to be covertly focused on the first letter, with attention dropping off slowly and smoothly to the right, and rapidly to the left (so as not to process a previous word). This would produce serial read-out of the letters, which would support learning of grapheme/phoneme correspondences. When such learning then sub-lexically produces the correct phonological representation, this would reinforce the specialized visual processing. Thus we propose a bootstrapping process – serial visual encoding allows phonological learning, which reinforces the serial visual encoding. Over time and experience, the visual system then learns to create the activation gradient in a bottom-up manner, corresponding to the proposed process of locational-gradient formation.

We propose that the failure to learn this specialized visual processing is a causal factor in dyslexia. This proposal is somewhat similar to those of Cestnick and Coltheart (1999) and Vidysagar (1999). They proposed that serial allocation of attention across a word is necessary in skilled reading. In contrast, we propose that allocation of a static attention gradient is necessary in learning to read. This activation gradient interacts with an innate
serial/oscillatory mechanism to produce sequential letter activations. In skilled reading, this activation gradient is formed automatically in a bottom-up manner. Thus we do not propose that skilled reading relies on serial visual attention. A failure to learn this specialized visual processing could stem directly from a visual deficit, or from decreased connectivity between auditory and visual areas.

If a budding reader cannot covertly focus attention away from fixation, s/he will not be able to produce a serial encoding when fixating near the center of the string. Indeed, a recent study showed that the one visual deficit that a group of eight adult English-speaking dyslexics all had in common was an inability to focus attention away from fixation (Buchholz & Aimola Davies, 2006). Why then would such a reader not simply fixate on the initial letter? It may be the case that inability to covertly focus attention is accompanied by an inability to distribute attention asymmetrically. That is, allocation of attention may be limited to widening or narrowing a symmetric window around fixation. Under such a limitation when reading text, fixation on the first letter of the current word would obligatorily induce a read out of a previous word (from right to left). In order to limit attention to the current word, fixation would then have to occur within that word.

If visual attention is normal, but connectivity from visual to auditory areas is reduced (Deutsch et al., 2005), serial visual processing will remain possible, but may not produce rapid learning of grapheme-phoneme correspondences. Therefore, serial visual processing may not be sufficiently reinforced, causing the bootstrapping to fail. Learning to read English, with its low consistency between spelling and sound, would be particularly challenging problem under such conditions. The lack of transparency would
make it harder to learn grapheme-phoneme correspondences in the first place, while a correctly learned mapping would be less likely to yield the correct phonological encoding and thereby reinforce the serial visual processing.

Using reinforcement-learning simulations (unpublished), Whitney has validated these basic scenarios. For high grapheme/phoneme consistency, rapid learning on connections between visual and auditory representations, and an initially moderately low probability of focusing attention on the first letter (around 0.2), the system learned the grapheme/phoneme correspondences and converged to the solution of consistently directing attention to the first letter (i.e., this probability became 1.0). This simulates the proposed bootstrapping process. Reduction of any one of these factors below a certain level causes the system to fail settle to this solution. This illustrates how an attention deficit could produce a failure to learn grapheme-phoneme mappings, without a core phonological deficit. Similarly, a reduced cross-area learning rate could result in abnormal visual processing of letter strings, without a visual deficit. As would be expected, if two factors are reduced concurrently, smaller reductions in each factor will induce failure. For example, a reduced cross-area learning rate could produce failure in a somewhat inconsistent language, yet sustain convergence in highly consistent language.

In contrast to the SERIOL model, the visual processing of letter strings in the parallel models is the same as for objects in general. Hence, these models cannot offer insights into visual deficits in dyslexia. The SERIOL model offers a comprehensive account of experimental results in normal visual word recognition and a novel approach to understanding visual processing in dyslexia.
References


Table 1

Mean Reaction Times (with Standard Deviations) in ms for targets as a function of prime condition and type of previous trial.

<table>
<thead>
<tr>
<th></th>
<th>Absent</th>
<th>Cross- position</th>
<th>Same- position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foil</td>
<td>447 (32)</td>
<td>448 (25)</td>
<td>450 (32)</td>
</tr>
<tr>
<td>Target</td>
<td>456 (37)</td>
<td>448 (31)</td>
<td>444 (35)</td>
</tr>
</tbody>
</table>
Table 2

Facilitation by prime type (RT in absent condition – RT in prime condition) in ms for each subject, broken down by type of previous trial.

<table>
<thead>
<tr>
<th>Previous trial = FOIL</th>
<th>C.W.</th>
<th>A.S.</th>
<th>M.H.</th>
<th>G.W.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross (312)</td>
<td>5</td>
<td>8</td>
<td>-8</td>
<td>-8</td>
</tr>
<tr>
<td>Same (123)</td>
<td>0</td>
<td>-6</td>
<td>0</td>
<td>-5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Previous trial = TARGET</th>
<th>C.W.</th>
<th>A.S.</th>
<th>M.H.</th>
<th>G.W.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross (312)</td>
<td>14</td>
<td>9</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Same (123)</td>
<td>15</td>
<td>6</td>
<td>18</td>
<td>5</td>
</tr>
</tbody>
</table>
Figure Captions

Figure 1: Overview of the three models under consideration. Units in the SERIOL model have graded activations, while units in the other models are on/off.

Figure 2: Illustration of the formation of the locational gradient at the feature level for the centrally-fixated stimulus *castle*. The vertical axis represents activation level, and the horizontal axis represents retinotopic space, with fixation at the origin. The processing is comprised of three transformations. For clarity, these transformations are shown as occurring sequentially, although they would actually occur interactively. In each frame, the boldface letters represent the activation pattern prior to the illustrated transformation, and the italics represent the result of the transformation. Bottom frame: the bold letters correspond to the acuity gradient. Stronger RH excitation brings the first letter (as well as the other RH letters) to a higher activation level. Middle frame: the first letter inhibits the second letter, and the first two letters inhibit the third. Thus RH, left-to-right inhibition inverts the acuity gradient. Top frame: cross-hemispheric inhibition “joins” the hemispheric gradients. The final result (bold letters in LVF/RH and italicized letters in RVF/LH) is a monotonically decreasing activation gradient.

Figure 3: Summary of experimental results from Peressotti and Grainger (1995). For same-position and cross-position primes (123 and 312), the bars represent presentation durations of 33, 50, and 67 ms, from left to right. For transposition (132, 321, 213) and replacement (d23, 1d3, 12d) primes, durations were 50 ms. Facilitation amounts for the
first three groups (same-position, cross-position, and transposition primes) directly give the experimental findings, which were obtained under a go/no-go procedure. The replacement primes were run under a yes/no procedure, which gave reduced facilitation overall. In order to compare replacement primes to the other conditions, their facilitations were scaled based on the results for transposition primes, which were also run under the yes/no procedure. To produce the scaled values, we compared the amount of facilitation for transposition primes under the go/no-go and the yes/no procedures, and multiplied the actual facilitations for the replacement primes by this factor, 1.7.

**Figure 4:** Results of mathematical model of data from Peressotti and Grainger (1995). The fit is based on a combination of fast parallel retinotopic priming, and slower serial abstract priming, as explained in the text.

**Figure 5:** Experimental results for Absent, Cross-position and Same-position primes, broken down by the type of the previous trial.
Figure 1
Figure 2
Figure 3
Figure 4
Figure 5