Readers Are Parallel Processors Trends in Cognitive Sciences

Joshua Snell, Jonathan Grainger

CellPress REVIEWS

Opinion

Readers Are Parallel Processors

Joshua Snell [1,2,3,*] and Jonathan Grainger [1,2,3]

Reading research has long endorsed the view that words are processed strictly one by one. The primary empirical test of this notion is the search for effects from upcoming words on readers' eye movements during sentence reading. Here we argue that no conclusions can be drawn from the absence of such effects, and that the serial versus parallel processing debate cannot be resolved without treading beyond the methodological scope of tracking eye movements. Recent considerations of how the brain organizes linguistic input have sparked key predictions in- and outside the realm of text reading, with ensuing research revealing phenomena that complicate the serial processing perspective. A case is made for parallelism, along with new methods to infer the cognitive architecture driving reading.

Highlights

Whether words are processed serially or in parallel continues to be a major debate in reading research. In the past decade, many researchers have embraced the seriality assumption.

This Opinion article shows that empirical findings were incorrectly thought to falsify parallel processing and that the seriality assumption works only if one treats the word recognition process as a black box.

The newest model of text reading comprises true word recognition mechanisms, causing it to spark fresh predictions. Successful tests of these predictions cannot be harmonized with serial processing.

Reading research is ready for a paradigm shift, both methodologically (treading beyond the measurement of eye movements in sentence reading) and theoretically (abandoning serial processing in favor of parallel processing).

The Serial Processing Ideal

It is evident that when we read, we adopt a largely serial strategy: texts are processed from left to right and from the top downwards, with the eyes' fixation jumping from one word to the next. Should we cease to impose seriality on the linguistic plane that bombards our retina with so many words at once, we would squander the canonical order of words and so a key ingredient of linguistic communication. Moreover, simultaneous processing of multiple words might conceivably incite word-to-word interference; for instance, through confusion about which letters belong to which word. For these reasons we can safely claim that during text reading we should ideally be able to confine our **attention** (see Glossary) to single words. But to what extent can we live up to this ideal?

In the course towards a full understanding of the **reading system**, the question of whether readers process multiple words simultaneously has retained much prominence. Precisely a decade ago Reichle, Liversedge, Pollatsek, and Rayner expressed in this journal the opinion that parallel word processing is implausible [1]. At the time, it had already become clear that serial and parallel processing frameworks were equally capable of accounting for eye movement behavior during text reading [2–4]. Therefore, Reichle *et al.*'s argument for serial processing was not necessarily data driven (although later empirical work was deemed to be in support of serial processing; see the next section), but rather one of parsimony: a serial processing system recognizes words in the intended order and does not mix-up information across words, given that only one word is attended at any time. By contrast, a parallel processing system might recognize words out of order, and 'if one were to simultaneously activate orthographic units for two words, this would produce noisy output corresponding to neither word' ([1], see p. 117).

A decade later, we can but agree that a serial processing system would have the simplest time reading, yet recent research has yielded various reasons to believe that the reading process is in fact not so simple. In this Opinion article we show that the serial processing assumption holds only if the word recognition process is treated as a **black box** – which has been a key aspect of both serial and parallel models of eye movements in text reading. The latest theoretical campaigns, marked by the implementation of word recognition mechanisms in models of text

[1]Laboratoire de Psychologie Cognitive, Centre National de Recherche Scientifique, France
[2]Institute of Language, Communication and the Brain, France
[3]Aix-Marseille University, Marseille, France

*Correspondence:
Joshua.Snell@hotmail.com (J. Snell).

reading [5], have sparked new critical predictions, empirical tests of which are in support of parallel processing.

Below, we first discuss why serial processing has been endorsed and show why some behavioral observations initially taken as strong evidence for serial processing may not be that. We subsequently discuss several phenomena predicted by the latest parallel processing framework and observed in ensuing experiments, that are difficult to reconcile with the seriality assumption. Taking these together, it appears sensible that readers are parallel processors and that paradigms beyond the realm of eye movements in sentence reading will shape the next trend of reading research.

A Need for New Empirical Strategies

Whether the brain deals with language in a serial or parallel fashion has been debated across multiple domains of reading research (Box 1). With respect to eye movements in text reading, computational modeling has been one important endeavor. However, as serial and parallel processing models accounted for reading behavior equally well [2–4], it soon became clear that the debate could not be settled without empirical inquiry. Although divided by their premises, serial and parallel processing frameworks have in recent years been gauged on the basis of at least two shared rationales. The first of these is that if words were processed in parallel, the speed of recognizing a word should be influenced by the frequency of the following word. While some corpus studies have reported such effects [6,7], standard experimental designs have produced mixed results in this regard [8,9]. However, even from a parallel processing perspective it may be fairly logical that such effects are fleeting. Word recognition speed, insofar as it is determined by bottom-up processing of visual

Box 1. Various Editions of a 50-Year-Old Debate

Throughout reading research's history, the serial versus parallel debate has continued to resurface in various forms addressing different levels of processing. Here are some examples from the past 50 years, and our opinion with respect to the status of these debates today.

(1) Letter Processing

The beginning of contemporary research on single word reading pitted Forster's serial search model [46] against Morton's parallel activation model [47]. These models contrasted both in the way letter identities were processed (left-to-right scan of letters versus parallel scanning), and in the way letter identities made contact with lexical representations (serial search vs parallel activation). A consensus has emerged in favor of parallel letter processing [48,49] and the parallel activation of lexical representations [23–29].

(2) Word Processing

Parallel processing of two simultaneously and briefly presented words was investigated in the early 1980s. Post-cued report of one of the words was sometimes contaminated by letter migrations from the other word, and this was taken as evidence for parallel processing of orthographic information across the two words [50,51]. These results and conclusions nicely anticipated the more recent results discussed in the present Opinion article and provided confirmation of an earlier finding in favor of parallel processing of word identities in sentences [52].

(3) Ambiguity Resolution

Language is rife with ambiguity, and its impact on linguistic processing has been investigated at multiple levels, with the same issue at stake: are different interpretations processed in parallel or is one processed before the other? At the single word level, the results of priming experiments suggest that readers entertain both meanings of homographs such as 'bank' (river/financial institution) and weigh them as a function of their relative frequency and any biasing context that might be present [53,54]. At the sentence level, the classic serial 'garden path' model [55] was challenged by the finding that ambiguity can sometimes facilitate processing [56]. This finding provides support for parallel models of sentence parsing [57,58] (see [59] for further discussion and [60–62] for other parallel models).

Glossary

Attention: the direction to – or filtering of – information in our immediate environment. Whereas proponents of serial processing believe that readers strictly attend to one word at a time (hence suppressing surrounding words, preventing those words from activating brain regions involved in language processing), proponents of parallel processing believe that readers generally attend to multiple words simultaneously. Attention can be directed to a region of our visual field without directly looking at it. We may also to some degree attend to locations or objects (causing those locations or objects to be processed by various brain regions) without necessarily being aware of it.

Black box: in the context of modeling the brain, the 'black box' refers to those cognitive mechanisms that are not specified or modeled. For instance, prominent models of eye movement behavior in text reading considered words to be recognized after prespecified amounts of time, depending on nothing more than word length and frequency. The word recognition process was thus a black box, within which it was unclear how letter recognition would lead to word activation and what factors may cause the recognition process to go awry.

E-Z Reader: the leading serial model of eye movement behavior in text reading, authored by Reichle and colleagues [2].

Interactive activation: the principle that cognitive stages have a mutual influence on processing at each respective stage. For instance, the excitation of neural nodes encoding certain letters may lead to the excitation of multiple neural nodes encoding various words, but those word nodes could in turn provide additional activation or attenuation of letter nodes, depending on whether those letters do or do not belong in the word.

Lexical processing: the activation of (clusters of) neurons encoding whole word representations. The recognition of letters is thought to lead to the activation of designated entries in our mental dictionary or so-called lexicon.

OB1-reader: the latest theoretical framework of reading, which, through its employment of a widespread attentional distribution and a spatiotopic sentence-level representation, is inherently a parallel processing model [5]. It also distinguishes itself from input, would mostly depend on the amount of processing resources allocated to each word in the visual field (i.e., the attentional distribution). There is no reason why this distribution should be modulated by word frequency (but see the next section, where we discuss the existence of indirect top-down effects of the upcoming word's frequency on word recognition speed).

The second assumption has been that if words are processed in parallel, information should be **spatially integrated** across words, such that readers will more rapidly recognize certain word characteristics (e.g., letters, sounds, meaning) if those characteristics are shared with surrounding words [10–15]. As it turns out, readers do integrate letter information across words [10–15], but higher-level integration effects (e.g., the word 'dog' being recognized faster when followed by 'cat' than when followed by an unrelated word) have remained largely elusive [12–15]. This has led researchers to believe that parallel processing may occur at sublexical levels, but that lexical access nonetheless occurs serially.

But this conclusion is not warranted. Parallel processing may well proceed without integrating high-level information across words [5,14,15]. The rationale for this alternative scenario is simple: readers would have to be able to keep track of which meanings belong to which positions in the sequence, given that each word has a unique role in contributing to sentence comprehension. A parallel processing system would be fundamentally flawed if it allowed cross-word leakage of high-level information (for an explanation of why letter information is nonetheless integrated across words, see Box 2). In neurophysiological terms, inherent to parallel processing must be a lack of lateral connections among the neural nodes representing semantic concepts activated by different words in the sentence.

Naturally one should not ponder the plausibility of this alternative scenario without having a concrete means of testing it. And here transpires a challenge: if words are truly processed in parallel (i.e., without integrating high-level information across words), 'direct' measures of word recognition

Box 2. Lessons from Research on Parafoveal Processing

As most words are big enough to occupy the whole region of sharp, central vision (the fovea), parallel word processing would necessitate some words to be processed in surrounding regions of the visual field, the parafovea. Therefore, an important question has been to what extent parafoveal words are processed at all. This has been investigated with the gaze-contingent boundary paradigm [63], whereby display manipulations prevent target word visibility until the eyes move to the target's location; (note, however, that some studies have shown that readers sometimes binocularly foveate two words simultaneously [64], which should arguably bode chaos for a serial processing system). A consistent finding is that denying parafoveal preview leads to prolonged target viewing times [65–67]. Consequently there has long been consensus that sublexical (letter) processing occurs for upcoming words. Higher-level (e.g., semantic, syntactic) preview was long considered controversial [68,69], but now researchers agree that upcoming words are also processed to such an extent [68–72]. A lack of such effects would certainly have put parallelism into question. However, the fact that words are parafoveally processed does not contradict seriality: the serial notion is that when a fixated word is recognized, attention moves, ahead of the eyes, to the upcoming word where high-level processing may subsequently occur.

Influences from word $n + 1$ on n would provide stronger evidence for parallelism, as $n + 1$ could have an impact on n only if it were processed simultaneously. While such effects are not found at levels of semantic processing [12,14,15] (which, as we argue in this Opinion article, is for good reason), they are consistently found at the level of letter processing, with words being recognized faster when followed by an orthographically related word than an unrelated word [10–13,18].

If the system prevents cross-word leakage of high-level information, why is the recognition process nonetheless influenced by surrounding words' letters? The answer is that the brain comprises location-aspecific orthographic nodes that are activated by features across the perceptual span [27,73]. Proponents of serial processing have argued that parafoveal feature detectors exert an influence on these letter detectors preattentively [12]. However, it has since been established that parafoveal influences pertain not just to featural overlap but also to the relative position of letters (such that 'rock' is recognized faster in 'ro rock ck' than in 'or rock kc'), which indicates that **orthographic processing** (i.e., the encoding of the identities and positions of letters) occurs across multiple words in parallel [74,75]. It has further been shown that this process is driven by attention [18].

prominent text reading models by having implemented mechanisms of letter identity and position coding and word recognition.

Orthographic processing: the encoding of the identities and positions of letters in the visual field and the interface between parts of the brain representing those letters and parts of the brain representing whole words.

Reading system: the collection of all brain regions or cognitive components involved in reading.

Recognition threshold: the point at which a cluster of neurons encoding a specific letter, a specific word, or any other relevant structure (e.g., syllable, phoneme, grammatical structure, language) has gathered enough evidence to signal certainty about the presence of whatever unit it represents. Note that word recognition in the more general sense touches on the concept of awareness and is likely to be dependent on convergence of processing along the whole neural network involved rather than the firing of a specific cluster of neurons. Hence, a neural node representing a word may reach a recognition threshold, but the experience of recognizing a word is not a punctuate event.

Spatial integration of information: in the context of the present Opinion article, this refers to the joint impact of multiple words in view on certain cognitive stages involved in reading. Among the cognitive stages are letter processing, word processing, and the processing of sentence structure (syntax) and word meaning (semantics).

speed (e.g., word viewing times) as a function of semantic relationships cannot be used to test parallelism. There is a solution, however. While information from simultaneously processed words is not integrated at the lexicosemantic level, these words may nonetheless jointly impact different, artificially induced measures, such as response times in a task where readers make decisions about words.

This proved to be true. In experiments where readers made semantic or syntactic categorization decisions about central target words, response times were influenced by respectively the semantic or the syntactic congruency of flanking words, even though targets and flanking words were presented simultaneously for only 170 ms [14,15]. Crucially, this is shorter than the average time required to recognize a single word [16], indicating that the flanking words were processed during rather than after target processing. Additionally, while performance was better in the presence of congruent than incongruent flankers, performance was better still when the target was presented in isolation [15]. That the presence of task-irrelevant flankers impairs performance indicates that they inevitably draw processing resources away from the target. This is fairly surprising under the assumption of serial processing, which dictates that readers without exception successfully confine their attention to single words.

More direct measures of visual attention have similarly indicated that attention is distributed across multiple words, even when the reader's aim is to focus on single target words. Compelling evidence is provided by pupil size measurements along with manipulation of the brightness of flanking words. Prior research has shown that the pupil size is contingent on the brightness of covertly attended (i.e., without looking) locations in the visual field [17]. Applying this principle in the flanker paradigm, it was found that the pupil size is contingent on the brightness of the locations of flanking words (with the overall display brightness kept equal) [18], indicating that during word processing a portion of attention is allocated to surrounding words.

Proponents of serial processing may aim to address these findings in two ways. The first option is to argue that artificial tasks such as the flanker paradigm are not representative of natural reading and that therefore no inferences can be made about the reading system. However, this does not solve the most obvious challenge: how could readers so effectively focus their attention on single words during normal reading, when this appears to be so difficult in a simple setting where the surrounding words' potential for distraction is considerably more constrained (given the 170-ms presentation time) and, moreover, surrounding words are irrelevant to the task at hand? Clearly the flanker paradigm differs from natural reading, but to claim that these settings would engage different systems for processing letters and words and focusing attention raises more questions than it answers. Further, while in the classical Eriksen flanker task [19] (using nonlinguistic stimuli) attention was found to be biased to the left [20,21], in the flanker paradigm using words attention is biased to the right, analogous to sentence reading [22]. This suggests that visual word stimuli may to a considerable extent engage the system in 'real' reading, regardless of the paradigm being used.

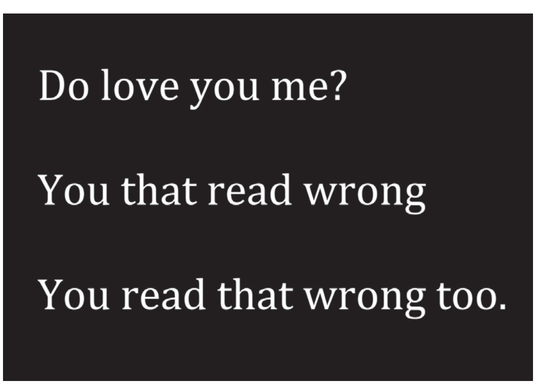
The second option is to acknowledge that visuospatial attention is indeed distributed across multiple words but that the system is somehow able to make lexical access proceed serially. However, this does not seem to hold either. First, under the assumption of serial lexical access, responses to target words in the flanker paradigm should not be influenced by the syntactic or semantic properties of flanking words, given that the flanking words will have disappeared before the target is recognized. Second, it is hard to see how processing resources could be allocated to surrounding words in the visual field without prompting activation of the neural nodes representing those words. Alternatively, should serial lexical access allow parallel word activation but permit only one word at a time to reach a **recognition threshold**? In any case, such conundrums demonstrate that the plausibility of serial lexical access cannot be contemplated so long as the word recognition process is treated as a black box.

The search for remedies has been hindered by a particular representation of the word recognition literature propelled by the article of Reichle *et al.* [1], who noted that '…**lexical processing** of multiple words is not consistent with any existing model of word identification' ([1], see p. 117). This claim is not true. Many models of word recognition apply **interactive-activation** principles whereby processing of a set of letters leads to the activation of multiple candidate word nodes, each of which in turn provides feedback activation/attenuation to letter detectors depending on whether those letters occur in the candidate word [23–29]. This back-and-forth process cycles until a candidate word reaches its recognition threshold. Ideally a single word – the correct one – would be recognized, but in principle multiple words can reach their threshold simultaneously: in the case of an unusual stimulus word (e.g., 'rook'), the reader may erroneously 'recognize' an orthographically related word of higher frequency ('rock') simply because the latter is activated faster and is therefore able to exert a stronger influence on letter detectors. This scenario is not merely a theoretical enterprise, but is evidenced by several key phenomena, such as neighborhood size and neighborhood frequency effects (orthographically similar words – e.g., 'hall', 'ball', 'bill', 'hill' – impact the ease with which single words are recognized [30–34]), embedded word effects (word recognition speed is influenced by the presence of embedded words – e.g., 'arm' and 'harm' in 'charm', [35,36]), and, more anecdotally, misreadings.

If attention is distributed across multiple words, and the reading system activates multiple lexical representations in parallel even when viewing a single word, is there anything still preventing the system from essentially being a parallel processing system? We shall now return to Reichle *et al.*'s main argument in favor of serial processing – that a parallel processing system would run into various problems – and show that the reading process can indeed go awry precisely because readers are parallel processors.

New Predictions of a Parallel Processing Framework

Concrete descriptions of how linguistic input is organized in the brain have been lacking in serial and parallel processing frameworks alike. Influential models have shown that when and where the eyes move can to a decent degree be predicted from the lengths and frequencies of words in the visual field [2–4], but these models have left unanswered important questions, such as how the swim through oceans of squiggly lines culminates in text comprehension, and under what circumstances this process might fail.



Trends in Cognitive Sciences

Figure 1. The Transposed-Word Phenomenon. The fact that many readers do not notice the transposed words should, according to Reichle *et al.* [1], not have been possible under the assumption of serial processing. Adapted from [38].

Aiming to solve this issue, a recent model of reading, **OB1-reader**, was equipped with mechanisms driving word recognition and mechanisms for associating activated words with locations in a sentence-level representation [5]. Its compliance with interactive-activation principles and the widespread attentional distribution as observed in the aforementioned studies naturally make OB1 a parallel processing model. Its theoretical crux is the interaction between location-independent activated words and a spatiotopic sentence-level representation in working memory. From the sentence level, feedback is provided to individual words based on top-down syntactic and semantic expectations; for instance, if an article is recognized at position 1, the reader may expect a noun or adjective at position 2, and a verb at position 3 would exert similar constraints. Activated words are mapped onto plausible locations. Crucially, this cognitive architecture provides an answer to Reichle *et al.*'s question of how a parallel processing system would handle the encoding of word order: the flexible, expectation-driven process of associating activated words with locations renders achronological word recognition nonproblematic.

Such theorizing sparks critical predictions that are unconceivable within a serial processing framework*. For instance, flexible, expectation-driven word position coding predicts that readers may confuse the order of words. Figure 1 provides a striking illustration that this can indeed be the case. Had words been processed in a serial left-to-right fashion, the transpositions in Figure 1 would be noticed without exception.

Although this provides a compelling argument against serial processing, evidence must of course be obtained in a formal experimental setting, and here one may realize that the transposed-word phenomenon cannot be experimentally established without again looking beyond behavioral patterns observed during normal sentence reading. A solution is found by letting readers make judgments about the grammaticality of sentences. In one study, it was established that readers have much more difficulty classifying a sequence as incorrect if that sequence can be 'corrected' through the transposition of two words (e.g., 'The ran dog slowly', which can be resolved into 'The dog ran slowly') compared with when the sequence cannot be corrected ('The was dog slowly') [37]. Readers' tendency to err more often in the case of a solvable sequence again shows that the process of word position coding is flexible and expectation driven. In a related study, the critical point of ungrammaticality (i.e., the point at which the reader can be certain that the sentence is incorrect) had no influence on the decision [38]. Had readers processed words in a serial left-to-right fashion, then earlier points of ungrammaticality should have produced faster decisions. It should also be noted that these patterns were observed while using infrequent sequences and transposing words of various syntactic categories [37,38]. Hence, word order confusion is not exclusively induced when using iconic sequences such as 'Do love you me'.

Another prediction of OB1-reader is that individual word processing is facilitated in a syntactically sound context compared with an incorrect context (a so-called sentence superiority effect). This prediction was successfully tested with the Rapid Parallel Visual Presentation (RPVP) paradigm. In this paradigm, readers were briefly presented four-word sequences, followed by masks (to prevent post-offset processing through sensory memory stores) and a cue for partial report at one of the word locations. Target recognition was considerably better in a correct sentence than when the same target was tested at the same location in a scrambled version of the same set of words [39]. Crucially, this sentence superiority effect was equal in size across all four

* See also the seminal work of Kennedy and Pynte [77] for an earlier appraisal of the importance of spatial information for reading, its role in the encoding of word order (especially when eye movements do not respect that order), and how the existence of a spatiotopic sentence-level representation *per se* is in conceptual discordance with serialism.

word locations, indicating that syntactic processing occurred across the whole sentence during its brief (200 ms) presentation time. Had words been processed serially, one would have expected larger effects at later word positions, given that these are constrained by a larger portion of preceding context.

Syntactic and semantic constraints are one way through which the frequency of surrounding words might indirectly impact recognition speed. The extent to which constraints are exerted is likely to depend on the speed of recognition of surrounding words and this in turn depends on those words' frequencies. Of course, in natural reading, variance in word-to-word constraint is much more subtle – and thus more difficult to observe – than in the RPVP paradigm, which blatantly contrasts syntactically legal versus illegal contexts. This would explain why frequency effects were observed in corpus studies [6,7] while being more elusive in standard experimental designs [11]. A reasonable prediction is that the sentence superiority effect should increase in size as the frequency of surrounding words increases.

One might reason that the sentence superiority effect observed in the RPVP paradigm was essentially a post-lexical memory effect; after all, prior research has shown that sentences (and their constituent words) are more easily remembered than random word sequences [40]. However, in a later study applying electroencephalography in the same paradigm, it was established that scrambled sentences evoked significantly increased negativity over anterior brain regions in the N400 window (associated with lexicosemantic access; [78]), with effects emerging as early as 270 ms post-stimulus onset [41]. Had the phenomenon been post-lexical in nature, the locus of effects should have been in the P600 window [42]. The timing of effects therefore strengthens the conception that the ongoing processing of individual words is affected by sentence structure.

How might these phenomena be accounted for by a serial processing framework? One possibility is that attention would rapidly dart back and forth across multiple words, such that while only one word is attended at any time, multiple words are nonetheless attended within the span of a single fixation. Indeed, in the leading serial processing model of reading, **E-Z Reader**, attention often moves to the next word ahead of the eyes when the fixated word has been processed to a certain extent [2]. The model would have to be adapted such that attention moves ahead quicker, and additionally attention should return to the fixated word in order for surrounding words to exert syntactic constraints. In our view, this is problematic for at least three reasons. Firstly, the total time spent attending each word would, by classic estimations of single word recognition speed, not be sufficient [16,28,29]. Second, in the programming of the next eye movement, E-Z Reader takes the attended location as the target location. This functional property of attention would be lost. The third problem is of a theoretical nature: serial processing would take on a ghostly quality with few testable consequences.

The final resort might again be to claim that the paradigms discussed here are not representative of natural reading, but invoking this argument would require an explanation of what cognitive systems drive behavior in these paradigms, if not the reading system [43,44]. The RPVP paradigm might encourage readers to adopt a parallel processing strategy (given that sentences are visible for only 200 ms), but this is certainly not the case in the grammaticality judgment task (where sentences are visible until the reader makes a decision) and in the flanker paradigm (where statically focusing on single targets benefits task performance).

Meanwhile, parallel processing smoothly handles phenomena in- and outside the realm of normal reading, which implies that it has the upper hand in terms of theoretical scope (see also Box 3 for answers to Reichle *et al.*'s Outstanding Questions [1]). The most natural conclusion, as we see it, is that readers are parallel processors.

Box 3. Outstanding Answers

To challenge parallel processing, Reichle *et al.* raised four 'Outstanding Questions' 10 years ago. We answer those questions here.

'If parallel processing causes some non-trivial proportion of words to be identified out of their canonical order, how are the comprehension problems that this would cause dealt with?'

Figure 1 shows that words may indeed be recognized out of order, and the fact that this can go unnoticed illustrates that our top-down expectations and flexibility in word position coding safeguard the coherence of the message being conveyed.

'How are the visual features that comprise the letters of a given word "bound" into a representation that is distinct from those of spatially adjacent words?'

They are not. As explained in Box 2, letter detectors are activated by information across the perceptual span, within the limits imposed by acuity and crowding. As a result, the recognition of words can be interfered with by letters from spatially adjacent words.

'How would attention-gradient models simultaneously explain the findings that denying preview of word *n* until it is fixated is rarely noticed and only modestly increases fixation durations on word *n*, but that the disappearance of word *n*+1 60 ms after word *n* is fixated is very disruptive?'

Research in the past 10 years has led to the consensus that denying preview of words until they are fixated is disruptive to the recognition process (Box 2). Thus, serial and parallel processing frameworks agree that processing of upcoming words starts prior to their fixation. The sudden disappearance of an upcoming word must therefore also be disruptive. Also note that sudden changes in the visual field trigger bottom-up attentional capture [76], so making a word disappear during the fixation on an adjacent word might not be a very appropriate manipulation to begin with.

'How are two words processed at the same time given existing models of word recognition?'

In almost all current models of word recognition, letters in the visual field activate multiple candidate word nodes [23–29]. The word recognition literature thus provides a solid argument for – rather than against – parallel processing.

Outstanding Questions

Given that letters from surrounding words can interfere with the recognition process, might some reading problems be caused by an attentional distribution that is too diffuse?

Is the seriality assumption tenable in light of the fact that simultaneous activation of multiple word nodes is required to account for all phenomena in the word recognition literature (e.g., neighborhood size and frequency effects, embedded word effects)?

As some languages give less priority to syntax (e.g., Chinese), might flexibility in word position coding – and the extent of parallel processing in general – differ across languages?

Might there be effective neuropsychological markers of parallel word processing, as measured with electroencephalography and fMRI?

Is knowledge of word order represented by a spatial code (e.g., a spatiotopic representation of the sentence in working memory) or a generic ordinal code, or both?

Concluding Remarks

In this Opinion article we have argued that the serial versus parallel processing debate cannot be resolved solely through observing eye movements during sentence reading. A natural reading setting is what one needs to describe reading behavior ('we move our eyes like such and so', the seminal work of Rayner [14] being a prime example) but not to infer the mechanisms driving reading behavior. By virtue of its parsimony and strictness, serialism has shaped clean and simple reasoning about a complex cognitive process. The aim to formulate and test more concrete accounts of how the brain organizes linguistic input, however, has prompted paradigms that appeal to specific parts of the system's core architecture. This endeavor has yielded sentence superiority effects, evidence that readers can confuse the order of words, evidence that attention is distributed across multiple words, and evidence that simultaneously processed words can jointly impact high-order decisions. Such phenomena complicate matters from a serial perspective. Through parallelism we may continue to effectively make sense of it all.

Importantly, the parallel processing framework discussed here generates additional predictions and perspectives for future research (see Outstanding Questions). For instance, we predict that, in sentence reading, word processing is influenced by the compatibility to the following word (such that 'dog' is recognized faster in the clause 'the dog barks' than in the clause 'the dog flies'). While such effects may be quite subtle (and possibly not observable in eye movement measures), one might employ electroencephalography with the onset of the interval of interest time-locked to the fixation on the target word. We predict effects analogous to the electroencephalographic patterns observed in the RPVP paradigm ([78]; see also [45] for evidence of semantic parallel word processing using fixation-related potentials in a dual word reading task).

The ultimate goal of this Opinion article is not to antagonize fellow researchers, but rather to amend the imbalance that has characterized an important debate – a debate whose quality may be improved through widening its theoretical scope (e.g., implementing mechanisms of word recognition and word position coding in models of text reading) as well as widening its methodological scope (e.g., employing paradigms beyond eye tracking in sentence reading). We have abundant reasons to believe that such a trend will ultimately lead the field to endorse parallel processing. Indeed, from a theoretical point of view, the observation that a strictly serial system is so easily lured into parallel processing upon the slightest deviation from text reading could be regarded as informative.

Acknowledgments

We are indebted to Ted Gibson for his advice, to editor Lindsey Drayton for her invaluable help in improving this Opinion article, and to Dr Richard Shillcock and two anonymous reviewers for their insightful comments. This work was funded by the European Research Council, grant ERC742141.

References

1. Reichle, E. *et al.* (2009) Encoding multiple words simultaneously in reading is implausible. *Trends Cogn. Sci.* 13, 115–119
2. Reichle, E. *et al.* (1998) Toward a model of eye movement control in reading. *Psychol. Rev.* 105, 125–157
3. Engbert, R. *et al.* (2005) SWIFT: a dynamical model of saccade generation during reading. *Psychol. Rev.* 112, 777–813
4. Reilly, R. and Radach, R. (2006) Some empirical tests of an interactive activation model of eye movement control in reading. *Cogn. Syst. Res.* 7, 34–55
5. Snell, J. *et al.* (2018) OB1-reader: a model of word recognition and eye movements in text reading. *Psychol. Rev.* 125, 969–984
6. Kennedy, A. and Pynte, J. (2005) Parafoveal-on-foveal effects in normal reading. *Vis. Res.* 45, 153–168
7. Kliegl, R. *et al.* (2006) Tracking the mind during reading: the influence of past, present and future words on fixation durations. *J. Exp. Psychol. Gen.* 135, 12–35
8. Angele, B. *et al.* (2016) Do successor effects in reading reflect lexical parafoveal processing? Evidence from corpus-based and experimental eye movement data. *J. Mem. Lang.* 88, 133–143
9. Brothers, T. *et al.* (2017) Looking back on reading ahead: no evidence for lexical parafoveal-on-foveal effects. *J. Mem. Lang.* 96, 9–22
10. Inhoff, A. *et al.* (2000) Allocation of visuospatial attention and saccade programming during reading. In *Reading as a Perceptual Process* (Kennedy, A., *et al.*, eds), Elsevier
11. Dare, N. and Shillcock, R. (2013) Serial and parallel processing in reading: investigating the effects of parafoveal orthographic information on nonisolated word recognition. *Q. J. Exp. Psychol.* 66, 417–428
12. Angele, B. *et al.* (2013) Parafoveal–foveal overlap can facilitate ongoing word identification during reading: evidence from eye movements. *J. Exp. Psychol. Hum. Percep. Perform.* 39, 526–538
13. Snell, J. *et al.* (2017) Integration of parafoveal orthographic information during foveal word reading: beyond the sub-lexical level? *Q. J. Exp. Psychol.* 70, 1984–1996
14. Snell, J. *et al.* (2017) Evidence for simultaneous syntactic processing of multiple words in reading. *PLoS One* 12, e0173720
15. Snell, J. *et al.* (2018) Parallel semantic processing in reading revisited: effects of translation equivalents in bilingual readers. *Lang. Cogn. Neurosci.* 33, 563–574
16. Rayner, K. (1998) Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422
17. Mathôt, S. *et al.* (2013) The pupillary response to light reflects the focus of covert visual attention. *PLoS One* 8, e78168
18. Snell, J. *et al.* (2018) Parallel graded attention in reading: a pupillometric study. *Sci. Rep.* 8, 3743
19. Eriksen, B. and Eriksen, C. (1974) Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* 16, 143–149
20. Harms, J. and Bundesen, C. (1983) Color segregation and selective attention in a nonsearch task. *Percept. Psychophys.* 33, 11–19
21. Hommel, B. (1995) Attentional scanning in the selection of central targets from multi-symbol strings. *Vis. Cogn.* 2, 119–144
22. Snell, J. and Grainger, J. (2018) Parallel word processing in the flanker paradigm has a rightward bias. *Atten. Percept. Psychophys.* 80, 1512–1519
23. McClelland, J. and Rumelhart, D. (1981) An interactive activation model of context effects in letter perception: part I. An account of basic findings. *Psychol. Rev.* 88, 375–407
24. Whitney, C. (2001) How the brain encodes the order of letters in a printed word: the SERIOL model and selective literature review. *Psychon. Bull. Rev.* 8, 221–243
25. Gomez, P. *et al.* (2008) The Overlap Model: a model of letter position coding. *Psychol. Rev.* 115, 577–600
26. Davis, C. (2010) The spatial coding model of visual word identification. *Psychol. Rev.* 117, 713–758
27. Grainger, J. and van Heuven, W. (2003) Modeling letter position coding in printed word perception. In *The Mental Lexicon* (Bonin, P., ed.), pp. 1–23, Nova Science
28. Grainger, J. (2008) Cracking the orthographic code: an introduction. *Lang. Cogn. Proc.* 23, 1–35
29. Carreiras, M. *et al.* (2014) The what, when, where and how of visual word recognition. *Trends Cogn. Sci.* 18, 90–98
30. Andrews, S. (1989) Frequency and neighborhood effects on lexical access: activation or search? *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 802–814
31. Grainger, J. and Segui, J. (1990) Neighborhood frequency effects in visual word recognition: a comparison of lexical decision and masked identification latencies. *Percept. Psychophys.* 47, 191–198
32. Davis, C. *et al.* (2009) Re(de)fining the orthographic neighborhood: the role of addition and deletion neighbors in lexical decision and reading. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1550–1570
33. de Moor, W. and Brysbaert, M. (2000) Neighborhood-frequency effects when primes and targets are of different lengths. *Psychol. Res.* 63, 159–162
34. Pollatsek, A. *et al.* (1999) The effects of “neighborhood size” in reading and lexical decision. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 1142–1158
35. Bowers, J. *et al.* (2005) Automatic semantic activation of embedded words: Is there a ‘hat’ in ‘that’? *J. Mem. Lang.* 52, 131–143
36. Snell, J. *et al.* (2018) A word on words in words: how do embedded words affect reading? *J. Cogn.* 1, 40
37. Mirault, J. *et al.* (2018) You that read wrong again! A transposed-word effect in grammaticality judgments. *Psychol. Sci.* 29, 1922–1929
38. Snell, J. and Grainger, J. (2019) Word position coding in reading is noisy. *Psychon. Bull. Rev.* 26, 609–615
39. Snell, J. and Grainger, J. (2017) The sentence superiority effect revisited. *Cognition* 168, 217–221

40. Baddeley, A. *et al.* (2009) Working memory and binding in sentence recall. *J. Mem. Lang.* 61, 438–456
41. Grainger, J. and Holcomb, P. (2009) Watching the word go by: on the time course of component processes in visual word recognition. *Lang. Linguist. Compass* 3, 128–156
42. Kuperberg, G. (2007) Neural mechanisms of language comprehension: challenges to syntax. *Brain Res.* 1146, 23–49
43. Grainger, J. and Jacobs, A. (1996) Orthographic processing in visual word recognition: a multiple read-out model. *Psychol. Rev.* 103, 518–565
44. Grainger, J. (2003) Moving eyes and reading words: how can a computational model combine the two? In *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (Hyönä, J., *et al*., eds), pp. 457–470, Elsevier
45. López-Peréz, P. *et al.* (2016) Semantic parafoveal-on-foveal effects and preview benefits in reading: evidence from fixation related potentials. *Brain Lang.* 162, 29–34
46. Forster, K. (1976) Accessing the mental lexicon. In *New Approaches to Language Mechanisms* (Wales, R. and Walker, E., eds), pp. 257–287, North Holland
47. Morton, J. (1969) Interaction of information in word recognition. *Psychol. Rev.* 76, 165–178
48. Adelman, J. *et al.* (2010) Letters in words are read simultaneously, not in left-to-right sequence. *Psychol. Sci.* 21, 1799–1801
49. Dehaene, S. *et al.* (2005) The neural code for written words: a proposal. *Trends Cogn. Sci.* 9, 335–341
50. Mozer, M. (1983) Letter migration in word perception. *J. Exp. Psychol. Hum. Percept. Perform.* 9, 531–546
51. McClelland, J. and Mozer, M. (1986) Perceptual interactions in two-word displays: familiarity and similarity effects. *J. Exp. Psychol. Hum. Percept. Perform.* 12, 18–35
52. Jackson, M. and McClelland, J. (1975) Sensory and cognitive determinants of reading speed. *J. Verbal Learning Verbal Behav.* 14, 565–574
53. Seidenberg, M. *et al.* (1982) Automatic access of the meanings of ambiguous words in context: some limitations of knowledge-based processing. *Cogn. Psychol.* 14, 489–537
54. Burgess, C. and Simpson, G. (1988) Cerebral hemispheric mechanisms in the retrieval of ambiguous word meanings. *Brain Lang.* 33, 86–103
55. Frazier, L. (1987) Sentence processing: a tutorial review. In *Attention and Performance* (Coltheart, M., ed.), pp. 559–586, Lawrence Erlbaum Associates
56. Traxler, M. *et al.* (1998) Adjunct attachment is not a form of lexical ambiguity resolution. *J. Mem. Lang.* 39, 558–592
57. Logačev, P. and Vasishth, S. (2016) A multiple-channel model of task-dependent ambiguity resolution in sentence comprehension. *Cogn. Sci.* 40, 266–298
58. van Gompel, R.P.G. *et al.* (2000) Unrestricted race: a new model of syntactic ambiguity resolution. In *Reading as a Perceptual Process* (Kennedy, A., *et al*., eds), pp. 621–648, Elsevier
59. Gibson, E. and Pearlmutter, N. (2000) Distinguishing serial and parallel parsing. *J. Psycholinguist. Res.* 29, 231–240
60. MacDonald, M. *et al.* (1994) Lexical nature of syntactic ambiguity resolution. *Psychol. Rev.* 101, 676–703
61. Vosse, T. and Kempen, G. (2000) Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition* 75, 105–143
62. Levy, R. (2008) Expectation-based syntactic comprehension. *Cognition* 3, 1126–1177
63. Rayner, K. (1975) The perceptual span and peripheral cues in reading. *Cogn. Psychol.* 7, 65–81
64. Shillcock, R. *et al.* (2010) Binocular foveation in reading. *Atten. Percept. Psychophys.* 72, 2184–2203
65. Briihl, D. and Inhoff, A. (1995) Integrating information across fixations during reading. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 55–67
66. Kliegl, R. *et al.* (2007) Preview benefit and parafoveal-on-foveal effects from word *n*+2. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 1250–1255
67. Inhoff, A.W. and Rayner, K. (1980) Parafoveal word perception: a case against semantic preprocessing. *Percept. Psychophys.* 27, 457–464
68. Rayner, K. *et al.* (2014) Lack of semantic parafoveal preview revisited. *Psychon. Bull. Rev.* 21, 1067–1072
69. Hohenstein, S. and Kliegl, R. (2014) Semantic preview benefit during reading. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 166–190
70. Schotter, E. (2013) Synonyms provide semantic preview benefit in English. *J. Mem. Lang.* 69, 619–633
71. Veldre, A. and Andrews, S. (2016) Is semantic preview benefit due to relatedness or plausibility? *J. Exp. Psychol. Hum. Percept. Perform.* 42, 939–952
72. Yan, M. *et al.* (2009) Readers of Chinese extract semantic information from parafoveal words. *Psychon. Bull. Rev.* 16, 561–566
73. Grainger, J. *et al.* (2016) A vision of reading. *Trends Cogn. Sci.* 20, 171–179
74. Grainger, J. *et al.* (2014) Test of a model of multi-word reading: effects of parafoveal flanking letters on foveal word recognition. *Acta Psychol.* 146, 35–40
75. Snell, J. *et al.* (2018) Parafoveal letter-position coding in reading. *Mem. Cogn.* 46, 589–599
76. Theeuwes, J. (1994) Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 799–806
77. Kennedy, A. and Pynte, J. (2008) The consequences of violations to reading order: an eye movement analysis. *Vis. Res.* 48, 2309–2320
78. Wen, Y. *et al.* (2019) Parallel, cascaded, interactive processing of words during sentence reading. *Cognition* 189, 221–226