Review

The what, when, where, and how of visual word recognition

Manuel Carreiras[1,2], **Blair C. Armstrong**[1], **Manuel Perea**[1,3], **and Ram Frost**[1,4,5]

[1] Basque Center for Cognition, Brain, and Language, San Sebastian, Spain
[2] IKERBASQUE Basque Foundation for Science, San Sebastian, Spain
[3] Universitat de València, València, Spain
[4] The Hebrew University of Jerusalem, Jerusalem, Israel
[5] Haskins Laboratories, New Haven, CT, USA

A long-standing debate in reading research is whether printed words are perceived in a feedforward manner on the basis of orthographic information, with other representations such as semantics and phonology activated subsequently, or whether the system is fully interactive and feedback from these representations shapes early visual word recognition. We review recent evidence from behavioral, functional magnetic resonance imaging, electroencephalography, magnetoencephalography, and biologically plausible connectionist modeling approaches, focusing on how each approach provides insight into the temporal flow of information in the lexical system. We conclude that, consistent with interactive accounts, higher-order linguistic representations modulate early orthographic processing. We also discuss how biologically plausible interactive frameworks and coordinated empirical and computational work can advance theories of visual word recognition and other domains (e.g., object recognition).

The what, when, where, and how of visual word recognition

A viable theory of visual word recognition needs to specify 'what' the building blocks of a printed word are and describe 'how' they are processed and assembled to give rise to word identification. These central 'what' and 'how' questions have been the focus of research (and controversy) in cognitive science since its very beginning, and have traditionally been addressed by combining inventive experimental designs and reaction time (RT) measures (Box 1). More recently, the availability of techniques such as functional magnetic resonance imaging (fMRI) have provided new opportunities to ask precise 'where' questions, focusing on locating the neurocircuitry involved in recognizing printed words. Given the architectural constraints of the brain, 'where' information often tells us something important about 'what' types of representations are activated during visual word recognition and 'how' readers eventually recognize words [1–3].

However, a comprehensive account of how complex stimuli such as words are processed requires a detailed description of the temporal flow of information and elucidation of 'when' the internal representations of words (e.g., letters, syllables, morphemes, lexical entries) are activated. Figure 1 presents contrasting frameworks. In this respect, 'when' questions constrain any theory of 'how' by detailing the sequence of events from stimulus presentation to word recognition. In fact, one of the oldest debates in visual word recognition concerns the demarcation between bottom-up and top-down processing, asking whether or not the visual stimulus feeds into the lexical level in a predominantly hierarchical manner, wherein orthographic representations feed into higher-level linguistic representations, or whether higher-level linguistic information such as phonological and morphological structure exerts a top-down influence on visual orthographic processing relatively early (Box 2). Cognitive neuroscience has rekindled this debate through the introduction of techniques such as electroencephalography (EEG) and magnetoencephalography (MEG), which have the appropriate temporal resolution to track the time course of processing. Note, however, that the 'where', 'what', 'how', and 'when' questions are to a large extent interdependent. The human brain is generally constructed so that the trajectory of increased complexity, in terms of moving from relatively simple microfeature representations (e.g., the line segments in a letter) to complex, higher-order representations (e.g., a representation of the whole word form) is occipital-to-frontal, whereas the trajectory of high-level modulation is frontal-to-occipital. Because 'where' information is correlated with the flow of processing (early/simple or late/higher-order), locations of brain activations are often taken to support claims regarding the temporal order of processing. Here we discuss the potential danger of using evidence of 'where' to make inferences about 'when' (and 'how'), review the findings obtained using techniques with the appropriate temporal resolution for tracking the time course of printed word processing, and point to desirable cross-fertilization between behavioral data, neuroimaging techniques, and neurobiologically plausible computational

---

Corresponding author: Carreiras, M. (m.carreiras@bcbl.eu).

Keywords: visual word recognition; visual word form area; orthographic processing; neural connectivity; computational modeling; feedback versus feedforward information.

1364-6613/$ – see front matter

http://dx.doi.org/10.1016/j.tics.2013.11.005

Box 1. Measuring time courses in behavioral research

Although behavioral investigations are useful for understanding visual word recognition, these studies suffer from an inherent limitation: they only provide an end-state reflection of the state of processing via an indirect behavioral response (e.g., lexical decision time as signaled via a key press). Thus, these techniques do not provide direct insight into the internal temporal dynamics underlying 'how' different representations are activated. Moreover, these approaches run the risk of producing data that are contaminated by pre- and post-lexical processes (e.g., decision-making).

These limitations notwithstanding, techniques that provide relevant indirect insight into the time course of different processes have been developed that relate to the 'when' question regarding feedforward and feedback processes. In this context, the masked priming technique [75] deserves special consideration. In masked priming, a target word is preceded by a briefly presented masked priming stimulus (e.g., mln-melon). By manipulating the structural relationships between prime and target (e.g., at the orthographic, phonological, morphological, and other levels) for different exposure durations (e.g., typically between 10 and 60 ms), researchers have observed different time courses of processing for different properties of printed words (e.g., orthographic and phonological representations) [76].

The rationale behind this experimental approach is that the minimal prime duration required to obtain a specific priming effect reflects the time necessary for activation of that information (e.g., orthographic, phonological, morphological, or semantic information). Nonetheless, this procedure has limitations [77], such as a lack of ecological validity. A related and more ecologically valid technique is to present the words in the context of normal silent reading while the participants' eye movements are registered [78]. Of particular interest is the very early parafoveal preview benefit effect using the boundary technique, in which the relationship between a parafoveal preview and a target word is manipulated. Specifically, the parafoveal preview is replaced by the target word once the fixation crosses an invisible 'boundary' located next to the target word. Differences in fixation duration on the target word caused by different structural manipulations of the parafoveal preview reflect 'what' information was already processed in the parafovea (e.g., orthography and/or phonology and/or morphology) [79].

There is ample evidence that high-level information, such as phonological [80,81], morphological [82,83], and lexical information [84], influences very early aspects of the overall visual word recognition process. This evidence challenges the traditional claim of temporal and structural modularity, according to which printed words are principally identified on the basis of orthographic information alone in skilled readers (the underlying logic behind some researchers' concept of the VWFA), with phonological and semantic information retrieved subsequently [64,85].

models for the development of a mechanistically explicit theory of visual word recognition.

fMRI evidence suitable for 'where' but not for 'when'

Many fMRI studies have investigated the brain circuits that underlie reading. Two points on which this research converge is that the left hemisphere plays a major role in reading and the reading circuit consists of a network with two major pathways: (i) a dorsal pathway including the occipital, supramarginal, and angular gyri, and the premotor and pars opercularis in the inferior frontal cortex; and (ii) a ventral pathway that integrates the left fusiform, middle and anterior temporal, and the pars triangularis in the inferior frontal cortex [4]. This notwithstanding, there is still a heated debate regarding the characterization of directionality of flow of information in these pathways (i.e., 'when' and 'how'). Specifically, the literature is unsettled regarding the extent to which higher-level lexical representations that are not necessarily orthographic modulate the relatively early processing of orthographic information (Box 3).

One of the most relevant examples of such debates is the role of the left fusiform gyrus, the putative visual word form area (VWFA) [5,6]. From an anatomical processing

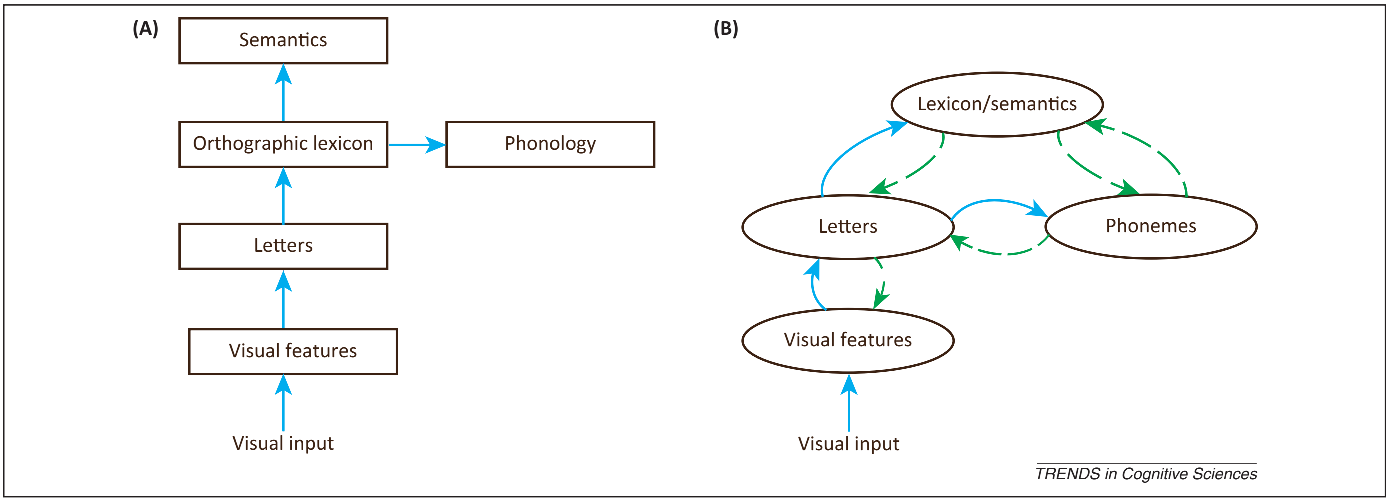

Figure 1. Core architectural and functional assumptions of temporally modular feedforward versus interactive models of visual word recognition. **(A)** According to temporally modular feedforward models, visual orthographic information is processed in a set of distinct, hierarchically organized processing stages, such that each stage (e.g., activation of letter and orthographic lexical representations) occurs in a strictly feedforward – and in the strongest form, sequential – fashion. Critically, additional non-visual orthographic representations (e.g., phonology, semantics) are not accessed until orthographic access is complete, and/or if accessed before that point, higher-level representations never feed back to influence the orthographic computation. **(B)** According to interactive activation models [59], visual information continuously cascades throughout the entire orthographic–phonological–lexical–semantic network. This enables partially resolved phonological and lexical–semantic representations (among others) to feed back and provide constraints on other (lower) levels of representation in the network such as orthography. Note that additional intermediate levels of representation (e.g., letter clusters) have been suppressed for simplicity in both schematics, and that these are just two examples of each type of network (e.g., other feedforward theories suggest a direct sublexical input to phonology but are nevertheless feedforward). Unbroken blue lines denote feedforward connections; broken green lines denote feedback connections.

Box 2. Structural and temporal modularity, interactivity, and orthographic processing

The main theoretical stand underlying the feedforward approach is that pure orthographic models have substantial descriptive adequacy to account for a large set of (mostly behavioral) data in visual word recognition [86]. One basic tenet in this feedforward view is that in principle, 'feedback cannot improve performance at either the lexical or prelexical level' [86,p. 306] and a number of well-known phenomena (e.g., word superiority effect) that have been traditionally attributed to top-down feedback can indeed be explained parsimoniously in a feedforward manner. According to this view, the game of processing printed words is largely played in the court of orthographic processing, such that a significant part of the recognition process is determined by considering the surface structural properties of the printed stimulus alone (i.e., letters, letter sequences). Interestingly, this position is consistent with the idea that the identification of visual forms in general and letter strings in particular can be achieved through a low-level visual pattern recognition system shared by humans and baboons [87–89].

The strongest version of the feedforward view postulates structural modularity, according to which orthographic processing is in principle non-penetrable by other linguistic dimensions. Moreover, processing within the orthographic system proceeds bottom-up from low-level features to full orthographic words. The weaker version assumes temporal modularity [90] and posits that the word recognition system is simply set so that the processing of printed words proceeds until an orthographic word unit is recognized; only when this is accomplished does the orthographic representation make contact with various other linguistic properties (e.g., phonology, morphology, semantics [64]). Note that this approach may allow for top-down interactivity as long as it is constrained to occur after initial orthographic coding is complete.

The contrasting approach argues for full interactivity between lower- and higher-order representations at all processing levels. Here, the demarcation line beyond 'when' and 'where' 'perceptual orthographic' processing ends and 'linguistic' processing begins is blurred [72]. According to this view, high-level linguistic considerations that are not purely orthographic (e.g., how some letters correlate with phonology and meaning, and how letter clusters are constrained by lexical, morphological, and phonological structure) shape the distributional properties of letters in a given language, and the word recognition system learns these features to enable efficient (i.e., fast and accurate) reading in that language. Language-specific retinal–perceptual learning effects (i.e., cross-linguistic difference in processing letters at different retinal eccentricities) [91] suggest that reading habits stemming from the overall structure of a language indeed affect the functional structure of early stages in the visual pathway, and are thus compatible with this view. For example, frequently encountered visual configurations result in perceptual learning that allows for rapid and efficient recognition of a word form, and these configurations are influenced by the correlation of orthography with phonology and meaning that are characteristic to a language [92]. Because different languages are characterized by different relations between orthography, phonology, and semantics (among other representations), it is argued that interactive models that allow for phonological, morphological, and semantic information to come into play early on are better accounts for the substantial cross-linguistic differences observed in early orthographic processing [85].

perspective (i.e., a 'where' distinction), this brain region is considered to be a relatively 'early' processing area. The left fusiform gyrus is activated more for words or pseudowords than for false fonts or consonant strings [7–10]. It is thus commonly accepted that the left fusiform is involved in orthographic processing. There is a controversy, however, regarding what specific information is represented in this brain region and how sensitive it is to top-down information. One theoretical position is that the VWFA is a prelexical hub, specific for written words, that computes and stores strictly visual and abstract prelexical orthographic representations in a primarily feedforward manner [8,11,12]. Another theoretical position, however, postulates that activation of the visual form area is modulated by higher-order linguistic properties of stimuli such as phonology, morphology, and semantics [13,14]. These two approaches provide very different views of reading: The former is compatible with the notion of feedforward temporal (and structural) modularity (Box 2), whereby reading is considered to rely on a sequence of consecutive brain areas sensitive to a hierarchy of orthographic representations (e.g., letters, letter clusters of increasing size) that culminates in recognition of a word. The latter considers reading as a fully interactive processing system whereby higher-level linguistic information that is not necessarily orthographic modulates early perceptual orthographic processing.

Whereas proponents of the feedforward approach have relied on the argument that VWFA activation reflects a stage of orthographic processing that is immune to phonological and semantic influences that come into play only later on [12,15,16], there is mounting evidence suggesting

Box 3. Future explorations of interactivity using fMRI and MEG

Several studies have shown intrinsic functional connections between Broca's area and ventral occipitotemporal regions [58,93–96] Anatomical connections between frontal and occipital regions through the superior longitudinal fasciculus and/or the inferior fronto-occipital fasciculus have also been documented [97–105]. Taken together, these findings provide a neurobiological platform for possible top-down effects from frontal areas, and thus offer interesting avenues for future investigations. Moreover, it is important to keep in mind that readers do not only activate the left fusiform and other regions of the ventral pathway when reading. They also activate the dorsal pathway, including regions such as the left supramarginal gyrus, the left superior temporal cortex, and the left inferior parietal cortex, and in particular the angular gyrus, which has a modulatory effect on the visual cortex [4,106]. Furthermore, there are other functional pathways in the reading circuit starting in the occipital cortex that do not necessarily involve the left fusiform [107]. The reading circuit includes not just one but multiple pathways from vision to higher-order temporal lobe language areas. Thus, the full scope of interactivity (or lack thereof) between regions spanning the different pathways of the reading circuit should be more fully established. Finally, it is important to note that the dorsal and ventral pathways are not modular systems that operate independently of each other, but exchange information during visual word recognition [108,109]. In fact, structural connectivity between regions belonging to each of the two pathways (e.g., the posterior inferior temporal regions, including the left fusiform, and the posterior superior temporal and inferior parietal regions, including the supramarginal gyrus) has been documented [110]. Moreover, functional connectivity between these regions has been shown in skilled readers [111] but not in dyslexic individuals [112]. Further studies are required to determine how brain regions falling along the ventral and dorsal pathways interact and cooperate during visual word recognition, and how these interactions relate to other similar processes such as object recognition.

that early print processing in the VWFA is modulated by high-level lexical information. For example, sensitivity to high-level variables such as lexical frequency has been observed in the left fusiform [17]. Furthermore, the VWFA was similarly activated when target words were preceded by masked printed-word primes or by masked pictures [18] (Box 1). Note that bidirectional information flow whereby higher-order levels of processing constrain feedforward assembly has also been proposed for object recognition [19–24]. Adopting the recycling hypothesis [12] (i.e., the neurocircuitry for visual object recognition is recycled to compute the representations necessary for human reading), simple parsimony considerations would lead to the assumption of similar principles regarding the flow of information for visual object and visual word recognition.

Despite the above evidence, the debate regarding whether processing of visual word recognition is feedforward or not is still as active as ever. This is because the critical distinction between the two conceptual approaches regarding 'how' information flows in the circuits is mostly temporal in nature ('early' vs 'late'). However, fMRI integrates processes over a relatively long period of time. Thus, although the timing of stimulus presentation can be well controlled (e.g., masked priming, fMRI adaptation), the temporal resolution of the blood-oxygen-level dependent (BOLD) response is too slow to unequivocally distinguish between activations that are feedforward versus feedback (i.e., the 'when' question) [25]. This leads us to the inherent advantage of MEG and EEG.

MEG and EEG: the dynamics of the time course

EEG and MEG are time-sensitive methods with a temporal resolution in the range of milliseconds. Combined with appropriate designs, they can reveal the temporal order of the neural processes involved in visual word recognition, tracing the time course from low-level visual perception to letter perception and word meaning. It is this time course that provides important evidence to adjudicate between different theoretical stances regarding the flow of information (feedforward vs feedback). Both techniques tap synchronized neuronal activity over time triggered by some cognitive event in the brain [26]. Unlike RTs, which give us the end result of processing in the system as a whole (and not specifically of lexical processing; Box 1), both MEG and EEG provide a continuous measure of the intermediate events that have led to the final response. MEG also provides some spatiotemporal constraints, allowing for some synthesis of 'when' and 'where' information.

One of the earlier markers of visual word recognition is a left lateralized N150/N170 response that differentiates orthographic stimuli such as words and pseudowords from other stimuli such as symbols [27–29]. Selective responses to letters in this time window have also been found in the inferior occipitotemporal cortex using intracranial recordings [30,31] and MEG [32,33], particularly for normal readers but not for dyslexic children [34–37]. Thus, it has been proposed that the left-lateralized N170 could be an automatic response related to typical visual word recognition, and that it could be associated with the activation found using fMRI in the VWFA left fusiform gyrus [38] (but see Rauschecker *et al.* [39] for bilateral effects). In fact, all this evidence supports the claim that at approximately 150 ms from stimulus onset, the visual system responds only to the frequency of letter combinations, and that lexical and phonological effects come into play much later [15,16,40]. As expounded below, however, higher-level linguistic information already exerts its influence at 100 ms (i.e., before 170 ms) from stimulus onset. For instance, it has been reported that early event-related potential (ERP) components in the range 100–200 ms are sensitive to lexical frequency [41–45]. Thus, from a simple time-scale perspective, an early marker of visual word recognition as revealed by ERP measures (but not fMRI measures) seems to be susceptible to modulation from higher-order lexical information.

Another early marker of reading is the N250, which was originally found to be sensitive to orthographic similarity in combined masked priming and EEG studies [46,47]. However, subsequent studies have shown that N250 is also modulated by lexical factors [48,49]. In particular, it was found that this ERP component is sensitive not just to letter identity but also to the phonological status of the letters, that is, whether letters are consonant or vowels [46,50–52]. For example, Carreiras and colleagues showed that masked subset priming of consonants (e.g., mln–melon) and masked full identity priming (e.g., melon–melon) do not significantly differ from each other in the N250 component, whereas masked vowel subset priming (e.g., aio–amigo) and masked identity priming (e.g., amigo–amigo) do [46]. Because consonants are more lexically constraining than vowels in predicting word identity [53], this effect demonstrates that top-down lexical information modulates the N250 component. Note that the same pattern of response is revealed in the later N400 component and in RTs in behavioral experiments. This suggests that accumulated lexical information (and/or lexical competition) that generates the masked prime has exerted its full impact by 250 ms from stimulus onset. In fact, the dissociation found between transposed-letter priming effects for word–word pairs (e.g., casual–causal) and for nonword–word pairs (e.g., barin–brain) in the N250 component [48] reinforces the hypothesis of high-order lexical–semantic information constraining orthographic form-level processing in the N250.

Consistent with sustained and early interactive coactivation of a network of sites contributing to reading, Thesen and colleagues found strong phase-locking from 170 to 400 ms between the left fusiform and more anterior language areas when comparing words versus false fonts using MEG and intracranial recording [54]. Other recent reports of very early neurobiological responses to phonological information in anterior areas are also consistent with a top-down flow of information during visual word recognition. Using MEG in a masked priming paradigm, Wheat and colleagues observed stronger responses to pseudohomophones than to orthographic control primes within 100 ms of target word onset in a cluster that included the left inferior frontal gyrus (pars opercularis) and the precentral gyrus [55–57]. Note that a parallel pattern of activation found in the middle occipital gyrus suggests that these regions could oscillate together during visual word recognition at a very early stage. Thus, the inferior frontal gyrus may exert feedback control on regions involved in lower-level analysis of written words. In fact, a recent study provides evidence of top-down feedback from the inferior frontal gyrus to the left ventral occipitotemporal cortex via dynamic causal modeling of MEG data [58]. Specifically, the researchers found that words (as compared to false fonts) activated the left inferior frontal gyrus. More importantly, they showed that feedback connections from the inferior frontal gyrus to the left ventral occipitotemporal cortex within the first 200 ms provided the best fit for the data relative to a model with only feedforward connectivity between these regions. One possible explanation for this feedback is that the inferior frontal region sends higher-level information (e.g., phonology) to constrain the representations computed in the left fusiform. Alternatively (or complementarily), these two brain regions may be interacting bidirectionally as part of a constraint network with as-yet underspecified graded specialization across the different contributing brain regions. Although the specific representations and dynamics of the frontal–occipitotemporal areas remain to be elucidated, these data clearly challenge the notion of temporal and structural modularity in orthographic processing.

Biologically plausible connectionist modeling: a platform for advancing theories of visual word recognition

From the empirical data outlined above, it is clear that substantive theoretical advances will require an integrated understanding of the contributions of a large set of distributed representations stored in different brain regions that are accessed (at least initially) at different points in time as activity cascades throughout the brain. Connectionist modeling offers a mechanistic platform that is ideally suited for these investigations because it allows researchers to probe the ‘where’ and ‘when’ of visual word recognition and directly relate them to the questions of ‘what’ (representations) and ‘how’ (explicit computational processing mechanisms) working in concert to enable the perception of written words. Moreover, models allow researchers to explore the emergent properties of these systems and develop targeted empirical research agendas for the future.

The basic capacities of connectionist networks as outlined above were keenly demonstrated via the interactive activation model [59–61]. In this model, information from low-level visual feature detectors flowed bottom-up to a lexical representation of whole words, while simultaneously being able to flow top-down from higher levels of representation. This model can thus explain and generate predictions regarding top-down influences related to word reading, such as the word superiority effect (i.e., the processing advantage for letters embedded in words relative to isolated letters) [62].

From this work and the general mechanics of constraint satisfaction systems [63], as well as the presence of bidirectional connectivity between brain regions that process different aspects of word representation (e.g., letters, phonology, semantics), connectionist theories of visual word recognition will, by default, expect and generate some degree of top-down influence for maximally accurate word identification. A more critical concern is if these top-down influences are substantive enough to be theoretically significant and may not be dismissed for reasons of parsimony [64–67]. Proficient reading involves optimization of efficiency in addition to accuracy, that is, correct reading of words as quickly as possible. Thus, if a strictly feedforward system could, in principle, enable highly efficient word recognition, under what circumstances would the brain choose to pay the price of waiting for additional top-down constraints because of inadequacies in a strictly feedforward signal [68]?

Connectionist models offer several avenues for exploring these possibilities. One particularly important recent advance is the ability to incorporate additional neurobiological constraints into standard connectionist models (e.g., by specifying different subpopulations of inhibitory and excitatory neurons) to simulate electrophysiological and behavioral responses (B.C. Armstrong, Ph.D. thesis, Carnegie Mellon University, 2012) [69,70]. For example, Laszlo and Plaut showed how a model that instantiates these principles can generate and explain electrophysiological dynamics corresponding to the N400 ERP component [69]. In addition, they were able to advance the field by offering an account of an important discrepant finding between behavioral and electrophysiological approaches: why the N400 ERP component is not sensitive to the lexicality of the stimulus (e.g., words and pseudowords vs acronyms and illegal strings), whereas behavioral responses are. Specifically, they showed that the initial settling dynamics, during which the prominent deflection typically associated with the N400 ERP component was displayed, were primarily driven by the orthographic wordlikeness of the stimulus (e.g., in terms of its orthographic neighborhood). However, nonlinear settling dynamics in the network caused a change in these activation patterns later on in processing, such that valid lexical types (words and acronyms) were more active than nonwords (pseudowords and illegal strings), consistent with typical behavioral lexical decision data.

Laszlo and Armstrong further extended this work to account for how simple context effects (e.g., word repetition) modulate the N400 component associated with lexical–semantic access [70]. This was accomplished via incorporation of a neuron-specific fatigue mechanism so that recently fired neurons would not be able to fire at their maximum rate for a brief period of time. This resulted in a reduction in N400-component amplitudes for stimuli in the semantic representation, regardless of the lexical status of the character string input to the network. Moreover, they were recently able to generate specific predictions regarding the power-frequency spectra that should be evoked by words and nonwords (Laszlo and Armstrong, unpublished), data that are increasingly influential in establishing the causal links between which brain regions influence one another and the temporal order (i.e., ‘when’) in which this occurs (Figure 2) [58,70]. This work led to targeted insights into ‘what’ aspects of a word representation are modulated by related context. Furthermore, this neurally inspired account therefore presents an alternative

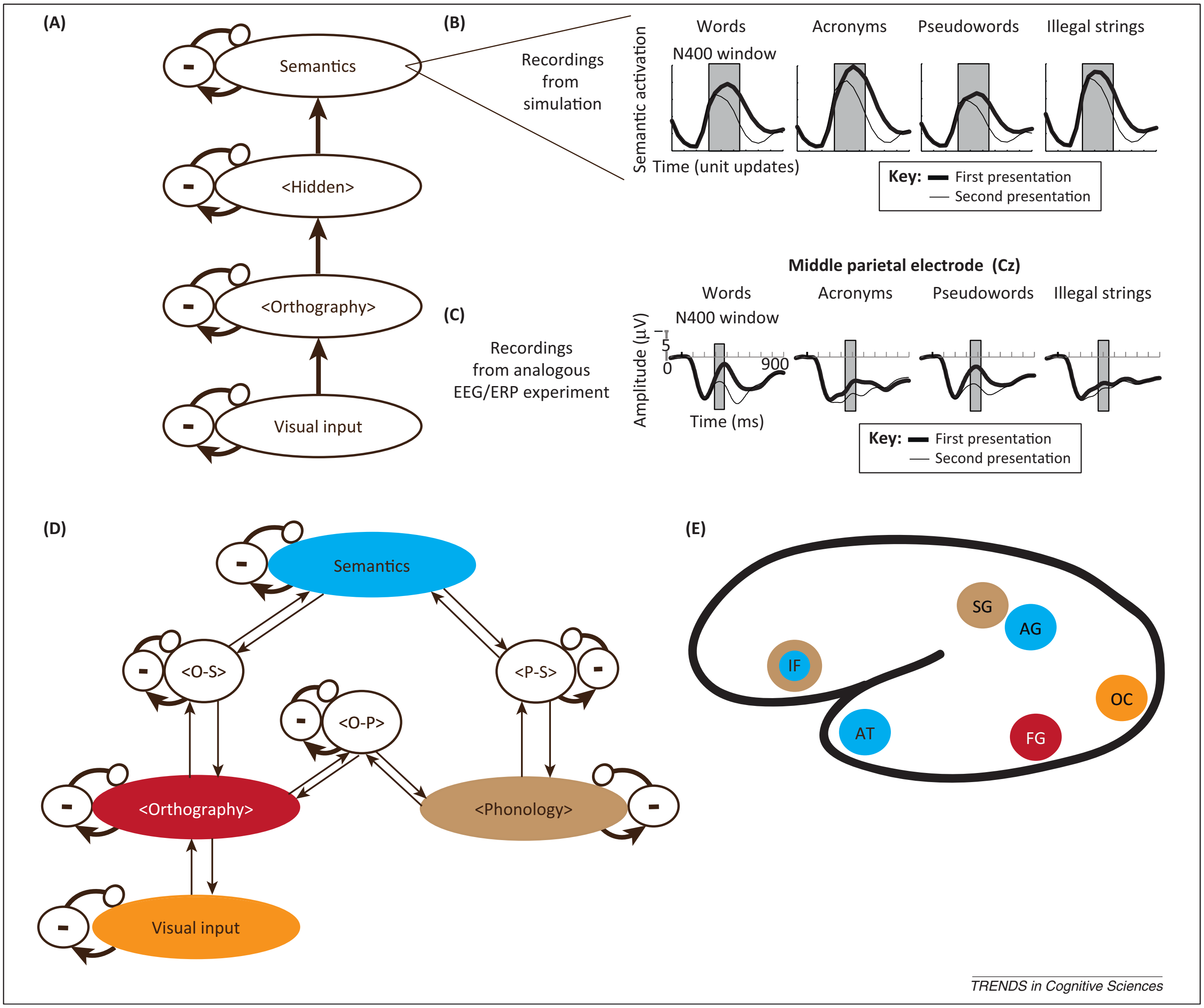

Figure 2. Integration of insights from more biologically plausible connectionist models and neuroimaging data. Recent connectionist models that use large pools of excitatory neurons and small pools of inhibitory neurons (here, inhibitory subpopulations are denoted by –; all other neurons are excitatory), as well as sparse/weak distal connections (thin arrows) and dense/strong local connections (thick arrows), and emergent hidden representations that are not explicitly specified by the modeler (denoted by < > around the name of the approximate representation that should emerge in that pool) have led to major advances towards explaining and predicting behavioral and neuroimaging data with a computationally explicit theory. For example, Laszlo and Plaut [69] and Laszlo and Armstrong [70] used the architecture illustrated in **(A)** to generate **(B)** simulated ERP components that explain empirical ERP data from **(C)** an analogous experiment involving lexical types including words (e.g., DOG), acronyms (e.g., DVD), pseudowords (e.g., DOD), and illegal strings (e.g., XFQ) both on their initial presentation (thick lines) and when the item was repeated (thin lines). Note that the units for the model data are arbitrary and are omitted; for the empirical data, the *x*-axis ranges from 0 to 900 ms and the negative direction is plotted upwards. All empirical and simulation data were drawn from Laszlo and Armstrong [70], used with permission. These event-related potential data were collected from an electrode placed on a middle parietal location. This work suggests that **(D)** in a more comprehensive model that also contains orthographic and semantic representations (here the <o-p>, <o-s>, and <p-s> labels denote intermediate pools of neurons that map between orthography, phonology, and semantics, respectively), it is in principle possible to study the activation dynamics and representations that emerge in **(E)** brain regions associated with different representations, of which a subset of the most critical regions are shown in a lateral cross-section of the left hemisphere. The color of the circles denotes the theoretical representations in the model that these regions might subserve. Abbreviations: IF, inferior frontal cortex; SG, supramarginal gyrus; AG, angular gyrus; AT, anterior temporal cortex; FG, fusiform gyrus (includes visual word form area, VWFA); OC, occipital cortex.

explanation for the long-accepted verbal account of N400 repetition effects [71], according to which reduced N400 amplitudes results from an (underspecified) facilitation mechanism rather than a fatigue mechanism.

With such explicit models in hand, it is possible to add or subtract different feedback connections and evaluate which of these models best captures empirical electrophysiological data such as ERP waveform amplitudes and power over time in as parsimonious a manner as possible. These models thus present an opportunity for an 'experimental' approach to theoretical development by allowing direct assessment of the impact of different theoretical assumptions. Moreover, by virtue of the domain-general nature of the framework, it is possible to naturally extend these principles to the study of other levels of representation. For instance, these principles can be readily applied to study ERP components associated with earlier processes related to visual word recognition (e.g., N170, N250) and determine 'when' and 'how' these representations shape other purportedly earlier processes.

Combined with a domain-general learning theory and recent advances in 'deep' neural networks [72], it has

Box 4. Outstanding questions and future directions

- How does anatomical and functional connectivity, and consequently the temporal flow of information, evolve from preliterate to skilled reading?
- What are the functional and anatomical differences underlying reading disabilities, and how can their understanding can help the implementation of remediation programs?
- How are connectivity and interactivity modulated by different languages having different scripts, different orthography-to-phonology relations, and different morphological systems?
- What do the constraints of possessing different orthographic, phonological, and semantic representations have on visual word recognition in the case of bilingual and multilingual readers?
- How can more biologically plausible computational models interact with empirical investigations to produce theories that are mechanistically explicit, comprehensive, and parsimonious?
- To what extent will a neurobiological theory of visual word recognition that considers the full patterns of brain connectivity and interactivity provide insights into domain-general mechanisms shared by other related abilities such as object recognition?

recently become possible to train networks with multiple intermediate 'hidden' layers whose input values or target outputs are not explicitly set by the experimenter. Consequently, researchers can step back from explicitly specifying the exact nature of the representations that are expected to mediate between a low-level visual form and higher-order lexical information. Instead, they can simply specify a more neutral input (e.g., a low-level visual input associated with a word in a retinotopically centered coding scheme) and observe how unsupervised learning mechanics can gradually extract higher-order information such as position-insensitive visual representations and intermediate representations of different grain sizes [73]. This approach allows empirical investigation of whether particular intermediate representations assumed to mediate between visual inputs and higher-order lexical outputs (e.g., bigrams, syllables) are in fact representations that will emerge during learning, and how these representations are constrained via feedforward and feedback constraints.

Taken together, biologically plausible connectionist models are showing strong initial promise in being able to provide detailed and explicit mechanistic accounts of 'what' and 'where' different types of representations are stored [72,73], 'when' they interact, and 'how' the processing mechanics of neurobiologically constrained computations operate the way they do. Moreover, these models can be used not only as 'process' models of the typical mature system but can also trace the development of lexical skills and top-down versus bottom-up interactivity [74], and can seamlessly facilitate studies of different types of reading impairments [61].

Concluding remarks

Old debates regarding the modularity – or lack thereof – of visual word recognition are back again with the advent of techniques such as fMRI, EEG, and MEG. However, rekindling of these debates via research using such techniques offers an opportunity for unique theoretical advance that was not possible via behavioral investigations alone, and also enables the investigation of important new questions (Box 4). No longer are researchers constrained to 'black box' theorizing regarding the internal mechanics of the brain that mediate between stimulus and response. Rather, these representations can now be monitored directly and used to motivate specific theoretical claims about the intermediate internal representations and processes that subserve visual word recognition. Like never before, it is therefore possible to achieve integrated theories of 'what', 'when', 'where', and 'how' visual words are represented and processed in the brain.

Acknowledgments

We acknowledge support of M.C. by the European Research Council (ERC-2011-ADG-295362) and by the Spanish Ministry of Economy and Competitiveness CONSOLIDER-INGENIO2010 CSD2008-00048 and PSI2012-31448; of B.C.A. by the Marie Curie program (PCIG13-2013-618429); of R.F. by the NICHD (RO1 HD067364); of R.F. and M.C. (PO1HD 01994); and of M.P. by the Spanish Ministry of Economy and Competitiveness (PSI2011-26924). We thank S. Laszlo for discussion of the modeling section.

References

1 Rogers, T.T. *et al.* (2004) Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychol. Rev.* 111, 205–235
2 Plaut, D.C. (2002) Graded modality-specific specialisation in semantics: a computational account of optic aphasia. *Cogn. Neuropsychol.* 19, 603–639
3 Plaut, D.C. and Behrmann, M. (2011) Complementary neural representations for faces and words: a computational exploration. *Cogn. Neuropsychol.* 28, 251–275
4 Price, C.J. (2012) A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62, 816–847
5 Cohen, L. *et al.* (2000) The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* 123, 291–307
6 Cohen, L. *et al.* (2002) Language-specific tuning of visual cortex? Functional properties of the visual word form area. *Brain* 125, 1054–1069
7 Baker, C.I. *et al.* (2007) Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9087–9092
8 Dehaene, S. *et al.* (2002) The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport* 13, 321–325
9 Glezer, L.S. *et al.* (2009) Evidence for highly selective neuronal tuning to whole words in the visual word form area. *Neuron* 62, 199–204
10 Binder, J.R. *et al.* (2006) Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage* 33, 739–748
11 Dehaene, S. *et al.* (2005) The neural code for written words: a proposal. *Trends Cogn. Sci.* 9, 335–341
12 Dehaene, S. and Cohen, L. (2011) The unique role of the visual word form area in reading. *Trends Cogn. Sci.* 15, 254–262
13 Price, C.J. and Devlin, J.T. (2003) The myth of the visual word form area. *Neuroimage* 19, 473–481
14 Price, C.J. and Devlin, J.T. (2011) The interactive account of ventral occipitotemporal contributions to reading. *Trends Cogn. Sci.* 15, 246–253
15 Simos, P.G. *et al.* (2002) Brain mechanisms for reading words and pseudowords: an integrated approach. *Cereb. Cortex* 12, 297–305
16 Solomyak, O. and Marantz, A. (2010) Evidence for early morphological decomposition in visual word recognition. *J. Cogn. Neurosci.* 22, 2042–2057
17 Kronbichler, M. *et al.* (2004) The visual word form area and the frequency with which words are encountered: evidence from a parametric fMRI study. *Neuroimage* 21, 946–953
18 Kherif, F. *et al.* (2011) Automatic top-down processing explains common left occipito-temporal responses to visual words and objects. *Cereb. Cortex* 21, 103–114
19 Bar, M. *et al.* (2006) Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci. U.S.A.* 103, 449–454

20 Bar, M. (2007) The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci.* 11, 280–289
21 Kveraga, K. *et al.* (2007) Magnocellular projections as the trigger of top-down facilitation in recognition. *J. Neurosci.* 27, 13232–13240
22 Gazzaley, A. *et al.* (2007) Functional interactions between prefrontal and visual association cortex contribute to top-down modulation of visual processing. *Cereb. Cortex* 17 (Suppl. 1), i125–i135
23 Zanto, T.P. *et al.* (2011) Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nat. Neurosci.* 14, 656–661
24 Bar, M. (2003) A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15, 600–609
25 Logothetis, N.K. (2008) What we can do and what we cannot do with fMRI. *Nature* 453, 869–878
26 Hämäläinen, M. *et al.* (1993) Magnetoencephalography – theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.* 65, 413
27 Dunabeitia, J.A. *et al.* (2012) Differential sensitivity of letters, numbers, and symbols to character transpositions. *J. Cogn. Neurosci.* 24, 1610–1624
28 Bentin, S. *et al.* (1999) ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *J. Cogn. Neurosci.* 11, 235–260
29 Maurer, U. *et al.* (2005) Fast, visual specialization for reading in English revealed by the topography of the N170 ERP response. *Behav. Brain Funct.* 1, 13
30 Allison, T. *et al.* (1994) Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cereb. Cortex* 4, 544–554
31 Nobre, A.C. *et al.* (1994) Word recognition in the human inferior temporal lobe. *Nature* 372, 260–263
32 Tarkiainen, A. *et al.* (1999) Dynamics of letter string perception in the human occipitotemporal cortex. *Brain* 122, 2119–2132
33 Tarkiainen, A. *et al.* (2002) Dynamics of visual feature analysis and object-level processing in face versus letter-string perception. *Brain* 125, 1125–1136
34 Helenius, P. *et al.* (1999) Dissociation of normal feature analysis and deficient processing of letter-strings in dyslexic adults. *Cereb. Cortex* 9, 476–483
35 Simos, P.G. *et al.* (2002) Dyslexia-specific brain activation profile becomes normal following successful remedial training. *Neurology* 58, 1203–1213
36 Simos, P.G. *et al.* (2005) Early development of neurophysiological processes involved in normal reading and reading disability: a magnetic source imaging study. *Neuropsychology* 19, 787–798
37 Simos, P.G. *et al.* (2007) Altering the brain circuits for reading through intervention: a magnetic source imaging study. *Neuropsychology* 21, 485–496
38 Brem, S. *et al.* (2010) Brain sensitivity to print emerges when children learn letter-speech sound correspondences. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7939–7944
39 Rauschecker, A.M. *et al.* (2012) Position sensitivity in the visual word form area. *Proc. Natl. Acad. Sci. U.S.A.* 109, E1568–E1577
40 Szwed, M. *et al.* (2012) Towards a universal neurobiological architecture for learning to read. *Behav. Brain Sci.* 35, 308
41 Assadollahi, R. and Pulvermuller, F. (2003) Early influences of word length and frequency: a group study using MEG. *Neuroreport* 14, 1183–1187
42 Carreiras, M. *et al.* (2005) Early event-related potential effects of syllabic processing during visual word recognition. *J. Cogn. Neurosci.* 17, 1803–1817
43 Dambacher, M. *et al.* (2006) Frequency and predictability effects on event-related potentials during reading. *Brain Res.* 1084, 89–103
44 Hauk, O. *et al.* (2006) The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage* 30, 1383–1400
45 Sereno, S. (2003) Measuring word recognition in reading: eye movements and event-related potentials. *Trends Cogn. Sci.* 7, 489–493
46 Carreiras, M. *et al.* (2009) Consonants and vowels contribute differently to visual word recognition: ERPs of relative position priming. *Cereb. Cortex* 19, 2659–2670
47 Holcomb, P.J. and Grainger, J. (2007) Exploring the temporal dynamics of visual word recognition in the masked repetition priming paradigm using event-related potentials. *Brain Res.* 1180, 39–58
48 Duñabeitia, J.A. *et al.* (2009) N250 effects for letter transpositions depend on lexicality: 'casual' or 'causal'? *Neuroreport* 20, 381–387
49 Massol, S. *et al.* (2011) When less is more: feedback, priming, and the pseudoword superiority effect. *Brain Res.* 1386, 153–164
50 Carreiras, M. *et al.* (2009) Are vowels and consonants processed differently? Event-related potential evidence with a delayed letter paradigm. *J. Cogn. Neurosci.* 21, 275–288
51 Carreiras, M. *et al.* (2007) ERP correlates of transposed-letter similarity effects: are consonants processed differently from vowels? *Neurosci. Lett.* 419, 219–224
52 Carreiras, M. *et al.* (2009) ERP correlates of transposed-letter priming effects: the role of vowels versus consonants. *Psychophysiology* 46, 34–42
53 Dunabeitia, J.A. and Carreiras, M. (2011) The relative position priming effect depends on whether letters are vowels or consonants. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 1143–1163
54 Thesen, T. *et al.* (2012) Sequential then interactive processing of letters and words in the left fusiform gyrus. *Nat. Commun.* 3, 1284
55 Cornelissen, P.L. *et al.* (2009) Activation of the left inferior frontal gyrus in the first 200 ms of reading: evidence from magnetoencephalography (MEG). *PLoS ONE* 4, e5359
56 Pammer, K. *et al.* (2004) Visual word recognition: the first half second. *Neuroimage* 22, 1819–1825
57 Wheat, K.L. *et al.* (2010) During visual word recognition, phonology is accessed within 100 ms and may be mediated by a speech production code: evidence from magnetoencephalography. *J. Neurosci.* 30, 5229–5233
58 Woodhead, Z.V. *et al.* (2012) Reading front to back: MEG evidence for early feedback effects during word recognition. *Cereb. Cortex* http://dx.doi.org/10.1093/cercor/bhs365
59 McClelland, J.L. and Rumelhart, D.E. (1981) An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychol. Rev.* 88, 375
60 Seidenberg, M.S. and McClelland, J.L. (1989) A distributed, developmental model of word recognition and naming. *Psychol. Rev.* 96, 523–568
61 Plaut, D.C. *et al.* (1996) Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychol. Rev.* 103, 56–115
62 Reicher, G.M. (1969) Perceptual recognition as a function of meaningfulness of stimulus material. *J. Exp. Psychol.* 81, 275–280
63 Ackley, D.H. *et al.* (1985) A learning algorithm for Boltzmann machines. *Cogn. Sci.* 9, 147–169
64 Davis, C.J. (2012) Developing a universal model of reading necessitates cracking the orthographic code. *Behav. Brain Sci.* 1, 21–22
65 Cohen, L. and Dehaene, S. (2004) Specialization within the ventral stream: the case for the visual word form area. *Neuroimage* 22, 466–476
66 Borowsky, R. and Besner, D. (1993) Visual word recognition: a multistage activation model. *J. Exp. Psychol. Learn. Mem. Cogn.* 19, 813–840
67 Borowsky, R. and Besner, D. (2006) Parallel distributed processing and lexical-semantic effects in visual word recognition: are a few stages necessary? *Psychol. Rev.* 113, 181–195
68 Plaut, D.C. and Shallice, T. (1993) Deep dyslexia: a case study of connectionist neuropsychology. *Cogn. Neuropsychol.* 10, 377–500
69 Laszlo, S. and Plaut, D.C. (2012) A neurally plausible parallel distributed processing model of event-related potential word reading data. *Brain Lang.* 120, 271–281
70 Laszlo, S. and Armstrong, B.C. (2013) Applying the dynamics of post-synaptic potentials to individual units in simulation of temporally extended ERP reading data. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (Knauff, M. *et al.*, eds), pp. 2826–2831, Cognitive Science Society
71 Rugg, M.D. (1985) The effects of semantic priming and work repetition on event-related potentials. *Psychophysiology* 22, 642–647
72 Hinton, G.E. and Salakhutdinov, R.R. (2006) Reducing the dimensionality of data with neural networks. *Science* 313, 504–507
73 Di Bono, M.G. and Zorzi, M. (2013) Deep generative learning of location-invariant visual word recognition. *Front. Psychol.* 4, 635

74 Harm, M.W. and Seidenberg, M.S. (1999) Phonology, reading acquisition, and dyslexia: insights from connectionist models. *Psychol. Rev.* 106, 491–528

75 Forster, K.I. and Davis, C. (1984) Repetition priming and frequency attenuation in lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 680

76 Ferrand, L. and Grainger, J. (1994) Effects of orthography are independent of phonology in masked form priming. *Q. J. Exp. Psychol.* 47, 365–382

77 Tzur, B. and Frost, R. (2007) SOA does not reveal the absolute time course of cognitive processing in fast priming experiments. *J. Mem. Lang.* 56, 321–335

78 Rayner, K. (1998) Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372

79 Bélanger, N.N. *et al.* (2013) Orthographic and phonological preview benefits: parafoveal processing in skilled and less-skilled deaf readers. *Q. J. Exp. Psychol.* 66, 2237–2252

80 Frost, R. (1998) Toward a strong phonological theory of visual word recognition: true issues and false trails. *Psychol. Bull.* 123, 71

81 Rastle, K. and Brysbaert, M. (2006) Masked phonological priming effects in English: are they real? Do they matter? *Cogn. Psychol.* 53, 97–145

82 Deutsch, A. *et al.* (2003) Early morphological effects in reading: evidence from parafoveal preview benefit in Hebrew. *Psychon. Bull. Rev.* 10, 415–422

83 Deutsch, A. *et al.* (2005) Morphological parafoveal preview benefit effects in reading: evidence from Hebrew. *Lang. Cogn. Process.* 20, 341–371

84 Dunabeitia, J.A. *et al.* (2009) There is no clam with coats in the calm coast: delimiting the transposed-letter priming effect. *Q. J. Exp. Psychol.* 62, 1930–1947

85 Frost, R. (2012) A universal approach to modeling visual word recognition and reading: not only possible, but also inevitable. *Behav. Brain Sci.* 35, 310–329

86 Norris, D. *et al.* (2000) Merging information in speech recognition: feedback is never necessary. *Behav. Brain Sci.* 23, 299–325

87 Grainger, J. *et al.* (2012) Orthographic processing in baboons (*Papio papio*). *Science* 336, 245–248

88 Ziegler, J.C. *et al.* (2013) What can we learn from humans about orthographic processing in monkeys? A reply to Frost and Keuleers (2013). *Psychol. Sci.* 24, 1870–1871

89 Frost, R. and Keuleers, E. (2013) What can we learn from monkeys about orthographic processing in humans? A Reply to Ziegler *et al.*. *Psychol. Sci.* 22, 1868–1869

90 Andrews, S. (2006) All about words: a lexicalist perspective on reading. In *From Ink Marks to Ideas: Current Issues in Lexical Processing* (Andrews, S., ed.), pp. 314–348, Psychology Press

91 Nazir, T.A. *et al.* (2004) Reading habits, perceptual learning, and recognition of printed words. *Brain Lang.* 88, 294–311

92 Frost, R. *et al.* (2012) Towards a universal model of reading. *Behav. Brain Sci.* 35, 263

93 Bitan, T. *et al.* (2007) Developmental changes in activation and effective connectivity in phonological processing. *Neuroimage* 38, 564–575

94 Bitan, T. *et al.* (2005) Shifts of effective connectivity within a language network during rhyming and spelling. *J. Neurosci.* 25, 5397–5403

95 Koyama, M.S. *et al.* (2010) Reading networks at rest. *Cereb. Cortex* 20, 2549–2559

96 Mechelli, A. *et al.* (2005) Dissociating reading processes on the basis of neuronal interactions. *J. Cogn. Neurosci.* 17, 1753–1765

97 Wakana, S. *et al.* (2004) Fiber tract-based atlas of human white matter anatomy. *Radiology* 230, 77–87

98 Barbas, H. (2000) Connections underlying the synthesis of cognition, memory, and emotion in primate prefrontal cortices. *Brain Res. Bull.* 52, 319–330

99 Bernal, B. and Altman, N. (2010) The connectivity of the superior longitudinal fasciculus: a tractography DTI study. *Magn. Reson. Imaging* 28, 217–225

100 Kringelbach, M.L. (2004) and E. T. Rolls, The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Prog. Neurobiol.* 72, 341–372

101 Makris, N. *et al.* (2005) Segmentation of subcomponents within the superior longitudinal fascicle in humans: a quantitative, *in vivo*, DT-MRI study. *Cereb. Cortex* 15, 854–869

102 Bürgel, U. *et al.* (2006) White matter fiber tracts of the human brain: three-dimensional mapping at microscopic resolution, topography and intersubject variability. *Neuroimage* 29, 1092–1105

103 Yeterian, E. and Pandya, D. (2010) Fiber pathways and cortical connections of preoccipital areas in rhesus monkeys. *J. Comp. Neurol.* 518, 3725–3751

104 Yeterian, E.H. *et al.* (2012) The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex* 48, 58–81

105 Yeatman, J.D. *et al.* (2013) Anatomy of the visual word form area: adjacent cortical circuits and long-range white matter connections. *Brain Lang.* 125, 146–155

106 Carreiras, M. *et al.* (2009) An anatomical signature for literacy. *Nature* 461, 983–986

107 Richardson, F.M. *et al.* (2011) Multiple routes from occipital to temporal cortices during reading. *J. Neurosci.* 31, 8239–8247

108 Rosazza, C. *et al.* (2009) Early involvement of dorsal and ventral pathways in visual word recognition: an ERP study. *Brain Res.* 1272, 32–44

109 Zachariou, V. *et al.* (2014) Ventral and dorsal visual stream contributions to the perception of object shape and object location. *J. Cogn. Neurosci.* 26, 189–209

110 Thiebaut de Schotten, M. *et al.* (2012) Learning to read improves the structure of the arcuate fasciculus. *Cereb. Cortex* http://dx.doi.org/10.1093/cercor/bhs383

111 Vogel, A.C. *et al.* (2012) The putative visual word form area is functionally connected to the dorsal attention network. *Cereb. Cortex* 22, 537–549

112 van der Mark, S. *et al.* (2011) The left occipitotemporal system in reading: disruption of focal fMRI connectivity to left inferior frontal and inferior parietal language areas in children with dyslexia. *Neuroimage* 54, 2426–2436