Triangulating the Reading Brain: Eye Movements, Computational Models, and EEG

Ronan G. Reilly

1 Introduction

Reading is a relatively recent cultural innovation, emerging in the last 5,000 years or so. It is an acquired skill involving the decoding of patterns of visual stimulation into a linguistic and, ultimately, a conceptual representation. Because it involves a considerable amount of learning and entails the interaction of several disparate brain areas, it is an ideal experimental domain to study the plasticity of the brain in a relatively pure form. As Huey (1908) observed in his seminal work *The Psychology and Pedagogy of Reading* first published just over 100 years ago, to understand reading fully would involve gaining a deep understanding of complex brain function. Huey's goal remains elusive to this day. What has changed is the availability of a powerful set of tools with which to explore the process: eye tracking, computational modelling, and electroencephalogram (EEG) recording. I will argue in this short chapter that it is only through the coordinated deployment of all three tools that we will get close to attaining Huey's goal.

2 Eye Movements in Reading

A reader's eyes move along a line of text in a sequence of fixations separated by jumps called saccades. Reading, therefore, takes the form of a series of "snapshots" during which textual information is acquired. During one such snapshot, the recognition of a single English word of average length takes about 100 milliseconds (Rayner and Pollatsek 1989). From the pioneering research of McConkie et al. (1988), it emerged that the effective target of a given eye movement is the centre of the to-be-fixated word. The goal of eye movements in reading appears to be to attain an optimal viewing position (OVP) on a word, which, if successful, facilitates rapid

R. G. Reilly (✉)
Department of Computer Science, National University of Ireland, Maynooth, Ireland
e-mail: ronan.reilly@nuim.ie

word recognition (O'Regan 1990). In reality, fixation locations are normally distributed with means somewhere between the beginning and the centre of the word. This latter position is referred to as the preferred viewing location (PVL; Rayner 1979). The eye frequently undershoots or overshoots the optimal position or even the word boundary with the consequence that extra fixations and/or movements back to previously read words may need to be made (Nuthmann et al. 2005).

All current accounts of word targeting in reading assume that the writing system serves up unambiguously delineated word "blobs" that act as targets for the saccade programming mechanism. This solves several problems at one fell swoop. There is no need, for example, to invoke a word segmentation algorithm to extract the individual words for targeting. However, it raises the question of what happens when the words in a writing system are not so conveniently delineated. What strategies do readers adopt when reading unspaced writing systems such as Thai or Chinese? This is still very much an open research question.

3   Computational Models

Over the last 20 years, there has been a burgeoning of computational models in the field of eye movements and reading. This growth has occurred partly as a result of the wealth of data generated by modern eye-tracking technology and partly because of the need to manage theory development for what turns out to be a complex interplay of cognitive, perceptual, and motor processes. Up until the early 1990s, the main types of theory in the field were informal, verbally specified ones. Morrison's (1984) model is a good example of this genre. While providing a plausible account of the phenomena of saccade targeting and word skipping in reading, Morrison's model still omitted crucial aspects of the process. For example, it could not account for spillover effects, where the processing on one fixation affects another. It was also hard to infer reliably testable predictions from the model because of the complex parallel interaction of different processes (e.g. word recognition and saccade preparation occur in parallel in the model). The management of this complexity clearly called out for computational modelling. Moreover, a computational model was required rather than a purely mathematical one, since it effectively involves the integration of a number of distinct mathematical models into a process-based account instantiated as a computer program. As Norris (2005) put it:

In research on word recognition, models don't just resolve debates over what theories predict, they are often the only way that even the theorists themselves can be sure what their theories predict. (Norris 2005, p. 333)

While he was referring specifically to visual word recognition (VWR) models, his remarks can be applied just as emphatically to models of eye movements in reading. Tellingly, the computational instantiation of the Morrison model as E-Z Reader (Reichle et al. 1998) led to several important changes in the original model's formulation arising from inconsistencies in the relative timing of several of the component processes. These inconsistencies would have been hard or impossible

to detect without the support and constraints provided by a computational implementation.

With the use of computational models now the norm in the field, a new issue arises regarding the precise relationship between model and motivating theory that it instantiates. Because of the complexity of models involved, the relationship can often be ambiguous and is frequently under-specified by the model designers. This can, in turn, lead to confusion about the testability of certain features of a model and the implications that this might have for its veridicality.

For example, in the case of the very successful interactive activation (IA) model of VWR (McClelland and Rumelhart 1981; Rumelhart and McClelland 1982), one of its features not especially deliberated over by the authors and by implication not attributed much theoretical weight was the input letter representation. The approach adopted involved having identical banks of features and letters replicated across the visual field. To most researchers, this was taken as a computational convenience and a way of skirting around what is still a hard problem in computational vision, namely position- and scale-invariant object recognition. Nonetheless, the letter representation approach adopted by the IA model became the focus of some criticism and empirical evaluation by other researchers (e.g. Mewhort and Johns 1988; Humphreys et al 1990; Davis and Bowers 2004).

The critique that certain letter-migration errors were precluded by the input's design was indeed valid and the results of Davis and Bowers' (2004) experiments were informative. Similarly, the studies of Humphreys et al. (1990) were on a much broader canvas than merely a critique of the input format for the IA model. However, one felt, to a large extent, that both critiques missed the point, since the core assumptions of the IA model were not heavily dependent on the precise nature of the letter representation used. A more central property of the whole family of IA models was that of interactivity between word, letter, and feature levels. Undermining the centrality of this property to the model's performance could be seen as much more damaging. In fact, Norris (1994) appeared to do just that with his Shortlist speech perception model. The Shortlist model, using a purely bottom-up architecture, demonstrated effects that had required top-down influences in the TRACE model of speech perception (a member of the IA family of models; McClelland and Elman 1986).

The case of the IA models is a good illustration both of how computational models can act as an important stimulus for research and of how there is little agreement about what constitutes a damaging critique. Notwithstanding Norris' (1994) findings, the IA framework of models has proved very productive in many varied cognitive modelling domains (e.g. Grainger and Jacobs 1998). The Glenmore model of reading (Reilly and Radach 2006), for example, is an IA model that can account within one mechanism for basic patterns of eye movement behaviour and accommodate a wide range of well-established empirical phenomena including parafoveal preview effects (Rayner and Pollatsek 1989).

The E-Z Reader model mentioned earlier (Reichle et al. 1998) can be regarded as a simulation of reading when higher-level linguistic processing is running smoothly (Reichle et al. 2003). The model aims to explain how lexical processing influences

the progress of the eyes through the text and provides a framework for understanding word identification, visual processing, attention, and oculomotor control that determines when and where the eyes move during reading (Reichle et al. 2003).

E-Z Reader works on the hypothesis that linguistic processing affects eye movements in two different ways. First, there is a relatively low-level linguistic process that keeps the eyes moving forward. Second, higher-level processing occurs in parallel with this low-level processing and is effective when the higher-level processing is having difficulty.

The SWIFT model (Engbert et al. 2005) embodies a number of features that set it apart from E-Z Reader. Most notably, the model assumes the parallel processing of several words in a given fixation, the number of such words being constrained by the extent of the perceptual span. In contrast, E-Z Reader assumes serial processing of words. Another distinguishing feature of SWIFT is that the triggering of a saccade is autonomous from word recognition. E-Z Reader, on the other hand, assumes word recognition or at least partial recognition of a word drives the reading process forward.

There is still, however, a gulf between the modelling architectures used to account for reading data (e.g. Glenmore, SWIFT, and E-Z Reader) and those necessary to account for the neural basis of reading. Indeed, the very existence of this gulf allows the proliferation of models that are, in my view, difficult to chose among on the basis of behavioural data alone. What are required are additional constraints from the neural substrate that the models purport to abstract from.

4   Electroencephalogram

Focussing on the neural foundations of reading is a significant challenge for a number of reasons: (a) Reading is an active process, so paradigms that restrict eye movements degrade the process under investigation; (b) reading involves coordinated activity in a variety of brain regions from the retina and primary visual cortex to integration areas to language areas; and (c) both inter-subject and intra-subject variability is high across various aspects of reading. Despite these difficulties, great progress has been made using oculomotor recording, EEG, and functional magnetic resonance imaging (fMRI). The primary limitation of fMRI is that its timescale is orders of magnitude slower than the timescales of interest in the reading brain.

EEGs are recordings of minute low-frequency electrical potentials on the surface of the skull produced by neural activity within the brain (typically 10–300 μV, 0.5–40 Hz). EEGs on the surface of the scalp reflect the synchronous activity of large populations of cortical neurons—of the order of 10,000 or so—with similarly aligned current flow. Variations in the magnitude of these potentials in the same spatial location are assumed to reflect underlying cortical processing activity. The primary limitation of EEG is that it is difficult to link the noisy EEG responses to particular aspects of brain function.

All current computational models of reading tend to be behavioural rather than neuroscientific, constructed on an empirical base derived mainly from studies of eye movement patterns. Unfortunately, several successful models can account more or less equally well for the same behavioural data. The most promising source of additional constraint comes from neuroscientific data, and the best current source of such data is cognitive electro-physiological studies. Such studies have the necessary temporal resolution to provide insights into the time course of reading (Barber and Kutas 2007). Curiously, to date there have been very few attempts to use this source of constraint in the development of reading models.

The benefits of … a dynamic interplay between computational models and empirical research are clearly evident in several computational models of VWR [visual word recognition] based largely on behavioral measures (reaction time and accuracy). By contrast, on the whole, there is no similar give-and-take between computational modelers and electrophysiological researchers, perhaps because computational models have been agnostic if not silent regarding the time courses of the various neurophysiological processes or the brain areas involved in VWR. (Barber and Kutas 2007, p. 100).

While Barber and Kutas' (2007) observations refer to VWR models, their comments are even more apposite for dynamic reading. Consequently, an overarching goal of co-registration research should be to help bridge the gap between current models of reading and the complex neural basis of the process. However, building that bridge will require the significant reworking of current models and the development of new EEG paradigms compatible with the model development enterprise.

Reading and Event-Related Potentials (ERPs) ERP analysis has the potential to provide us with an exquisitely precise tool for revealing the temporal dynamics of the component processes of reading. Figure 1 is a schematic representation of the locus, size, and reliability of ERP effects associated with different levels of analysis of the VWR (from Barber and Kutas 2007, Fig. 4). Note that P1 and P2 are the first two positive peaks; N400 is a robust negative potential peak occurring typically around 400 ms. LPC represent later positive components such as the P600 (Osterhout et al. 1994). The figures in boxes represent size of effects in milliseconds (i. e. the temporal responsivity of the component amplitude)—darker figures are for effects supported by more than one study. As can be seen from this figure, ERPs can be used to index the temporal stages of the subcomponents of the reading process. Note that it is not just peak amplitude that is the source of information about timing but also the point at which the ERPs diverge as a function of an experimental manipulation. There is still some debate about the precise timing and nature of word identification, since the N400 is rather late compared to timings derived from eye movement behaviour, such as fixation durations (Sereno and Rayner 2003; Kliegl et al. 2006).

Fixation-Related Potentials (FRPs) Although still very much at an exploratory stage, the co-registration of eye movements and EEG has been successfully employed by several research groups (e.g. Baccino et al. 2005; Hutzler et al. 2007; Dambacher and Kliegl 2007; Dimigen et al. 2011).
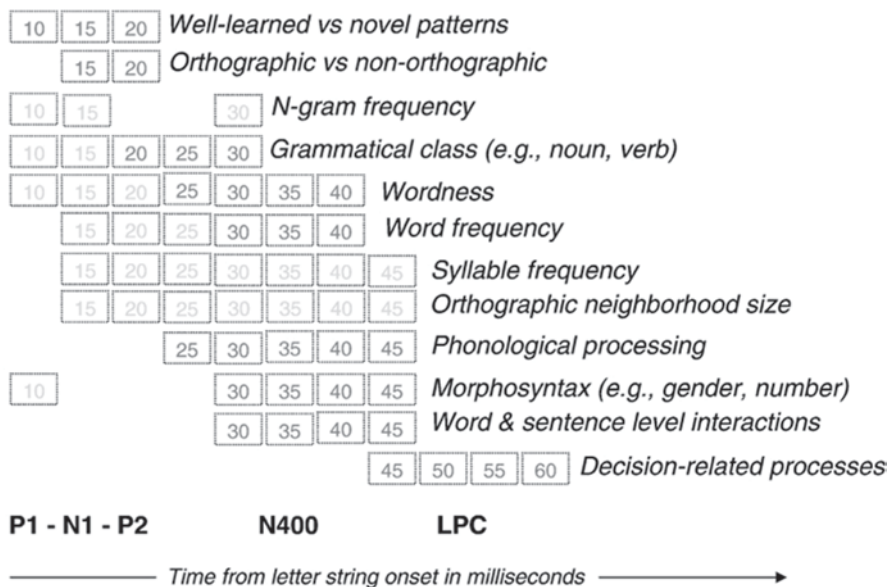
10	15	20					Well-learned vs novel patterns
	15	20					Orthographic vs non-orthographic
10	15			30			N-gram frequency
10	15	20	25	30			Grammatical class (e.g., noun, verb)
10	15	20	25	30	35	40	Wordness
	15	20	25	30	35	40	Word frequency
	15	20	25	30	35	40	45	Syllable frequency
	15	20	25	30	35	40	45	Orthographic neighborhood size
			25	30	35	40	45	Phonological processing
10				30	35	40	45	Morphosyntax (e.g., gender, number)
				30	35	40	45	Word & sentence level interactions
					45	50	55	60	Decision-related processes

P1 - N1 - P2   **N400**   **LPC**

————————— *Time from letter string onset in milliseconds* —————→

Fig. 1 Time course of effects in visual word recognition. (From Barber and Kutas 2007)

One of the disadvantages of permitting eye movements when recording EEGs is having to deal with artefacts from the movement of the eyes themselves. The eye is, in effect, a large dipole, any movement of which causes significant potential flows when viewed against the background of the much smaller scalp potentials. Nonetheless, the detection and removal of these artefacts is now relatively straightforward. For example, the use of independent component analysis (ICA) has been shown to be quite effective (Hutzler et al. 2007; Tang et al. 2002; Pearlmutter and Jaramillo 2003; Tang and Pearlmutter 2003; Henderson et al. 2013). Moreover, since the time-locking event is now the start of the fixation (hence, the term fixation-related potentials—FRPs—rather than ERPs), from the eye-tracking record we will know precisely when an eye movement has occurred.

The use of a more ecologically valid reading setting has many advantages. For example, in some reading studies involving a lexical decision task where the subject must press one of two buttons to indicate whether a word or non-word is being displayed, one can obtain P300 components that are merely associated with the need to generate a binary response (Kutas and van Petten 1994). These, in turn, can overlap with N400 responses, which are the usual focus of interest in lexical decision experiments. In contrast, by allowing free viewing and using the standard eye movement parameters of fixation duration and saccade extent, we get closer to the real reading process and avoid such procedural artefacts.

Naturally, the analysis of FRPs is not without its own significant challenges. Foremost among these is the problem of spillover of FRP components from one fixation to the next.
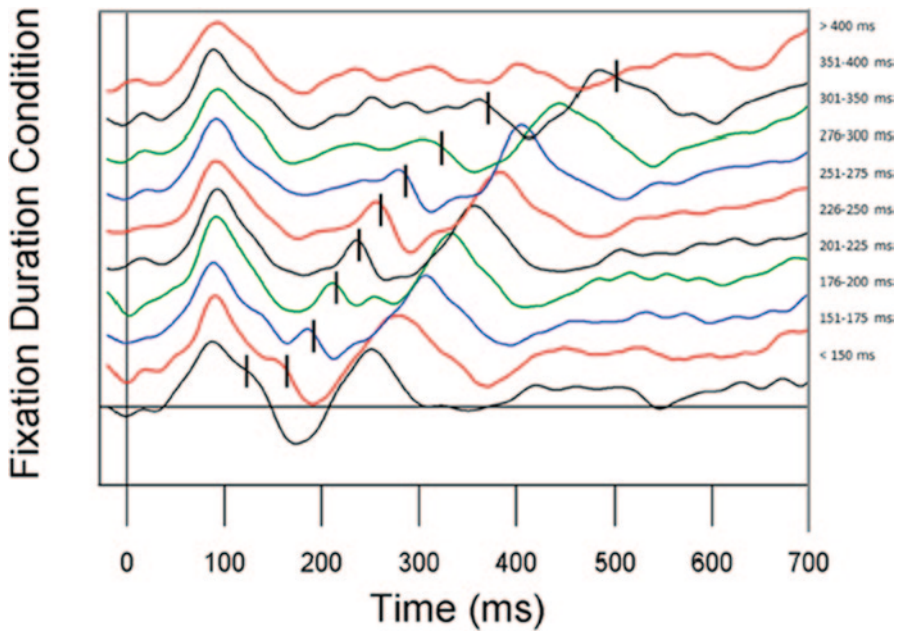
Fig. 2 Electroencephalogram (EEG) waveforms for fixations of different durations from Henderson et al. (2013). The *vertical bars* indicate fixation offset. Note that for short fixations, the P1 peak following fixation offset and associated with the new fixation can occur before some of the later components (e.g. N400) of the preceding fixations

Figure 2 from Henderson et al. (2013) gives a striking example of how spillover can make the interpretation of waveforms from natural reading problematic. In the case of, say, the 151–175-ms waveform in Fig. 2, it is impossible to distinguish the source of the negative inflection following the peak at P1. It could be the expected N1 for the second fixation, or it could have arisen from a possible N400 from the preceding fixation. A possible way to deal with this challenge is to provide context for the interpretation in the form of predictions from a computational model that would allow us to systematically unpack the various contributions to the final waveform.

5 A Synergistic Alliance

Two complementary challenges have been discussed in this chapter: (1) the need to determine which of a number of competing computational accounts of the reading process is the more plausible and (2) how to handle the complex spillover effects we find in co-registered EEG and eye movement data from natural reading experiments. However, by placing one challenge at the service of the other—in other words, use computational models to help disentangle the multiplexed EEG

waveforms, and use EEG data to ground the current generation of reading models—we may end up with a powerful and productive alliance.

Barber and Kutas (2007) made a similar appeal several years ago in the context of VWR models (2007):

> We suggest that it is time that computational modelers and neurophysiologists come together in practice and in theory to unravel the mysteries of reading (Barber and Kutas 2007, p. 119).

The pressure to combine forces is even more pressing in the case of the co-registration paradigm in reading, if only to make analysing the data more tractable.

A recent approach to neurally grounded modelling at the level off EEG-generating current flows has been that of dynamic causal modelling (DCM; Kiebel et al. 2008; Stephan et al. 2007). DCM aims to account for EEG data in terms of coupled neuronal groups and analyses how the topography of the couplings variously impact on the brain's response to different experimental conditions. Bayesian methods are then used to select the most likely candidate from among competing patterns of connectivity. This and approaches at a similar level of neural granularity are the next step forward in the computational modelling of reading.

References

Baccino, T., Manunta, Y. (2005). Eye-fixation-related potentials: Insight into parafoveal processing. *Journal of Psychophysiology*, *19*, 204–215.

Barber, H. A., & Kutas, M. (2007). Interplay between computational models and cognitive electrophysiology in visual word recognition. *Brain Research Reviews*, *53*, 98–123.

Dambacher, M., & Kliegl, R. (2007). Synchronizing timelines: Relations between fixation durations and N400 amplitudes during sentence reading. *Brain Research*, *1155*, 147–162.

Davis, C. J., & Bowers, S. J. (2004). What do letter migration errors reveal about letter position coding in visual word recognition? *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 923–941.

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Co-Registration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General*, *140*, 552–572.

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777–813.

Grainger, J., & Jacobs, A. (1998). *Localist connectionist approaches to human cognition*. Hillsdale: Erlbaum.

Henderson, J. M., Luke, S. G., Schmidt, J., & Richards, J. E. (2013). Co-registration of eye movements and event-related potentials in connected-text paragraph reading. *Frontiers in Systems Neuroscience*, *7*, 28. doi:10.3389/fnsys.2013.00028.

Huey, E. B. (1908). *The psychology and pedagogy of reading*. New York: Macmillan.

Humphreys, G. W., Evett, L. J., & Quinlan, P. T. (1990). Orthographic processing in visual word recognition. *Cognitive Psychology*, *22*, 517–560.

Hutzler, F., Braun, M., Vo, M., Engl, V., Hoffman, M., Dambacher, M., Leder, H., & Jacobs, A. M. (2007). *Brain Research*, *1172*, 124–129.

Kiebel, S. J., Garrido, M. I., Moran, R. J., & Friston, K. J. (2008). Dynamic causal modelling for EEG and MEG. *Cognitive Neurodynamics*, *2*, 121–136.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology:General*, *136*, 520–529.

Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 83–143). San Diego: Academic Press.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, *18*(1), 1–86.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological review*, *88*(5), 375.

McConkie, G., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words. *Vision Research*, *28*, 1107–1118.

Mewhort, D., & Johns, E. E. (1988). Some tests of the interactive activation model for word recognition. *Psychological Research Psychologische Forschung*, *50*, 135–147.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 667–682.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189–234.

Norris, D. (2005). How do computational models help us develop better theories? In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 331–346). Hillsdale, NJ: LEA.

Nuthmann, A., Engbert, R., & Kliegl, R. (2005). Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision Research*, *45*, 2201–2217.

O'Regan, J. K. (1990). Eye movements and reading. In E. Kowler (Ed.), *Reviews of Oculomotor Research, Vol. 4: Eye Movements and their Role in Visual and Cognitive Processes* (pp. 395–453).

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *28*, 786–803.

Pearlmutter, B. A., & Jaramillo, S. (2003). Progress in blind separation of magneto-encephalographic data. In Independent Component Analyses, Wavelets, and Neural Networks, volume 5102 of Procceding of the SPIE, pp. 129–34.

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, *8*, 21–30.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Boston: Prentice-Hall.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125–157.

Reichle, E., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: Accounting for initial fixation locations and refixations within words. *Vision Research*, *39*, 4403–4411.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E–Z reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*, 445–526.

Reilly, R., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research*, *7*, 34–55.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological review*, *89*(1), 60.

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*(11), 489–493.

Stephan, K. E., Harrison, L. M., Kiebel, S. J., David, O., Penny, W. D., & Friston, K. J. (2007). Dynamic causal models of neural system dynamics: Current state and future extensions. *Journal of Biosciences*, *32*, 129–144.

Tang, A. C., & Pearlmutter, B. A. (2003). Independent components of magneto-encephalography: Localization and single-trial response onset detection. In Z.-L. Lu & L. Kaufman (Eds.), *Magnetic source imaging of the human brain* (pp. 159–201). Lawrence Erlbaum Associates.

Tang, A. C., Pearlmutter, B. A., Malaszenko, N. A., Phung, D. B., & Reeb, B. C. (2002). Independent components of magnetoencephalography: Localization. *Neural Computation*, *14*(8), 1827–1858.