UNIVERSITY OF PITTSBURGH

Study and Detection of Mindless Reading

(Ph.D. Dissertation)

Tomasz Dominik Loboda[1]

Committee in charge:

Marek J. Druzdzel[1] *(chair), Erik D. Reichle*[2] *(co-chair), Peter Brusilovsky*[1]*, and Stephen C. Hirtle*[1]

March 17, 2014

[1]School of Information Sciences, University of Pittsburgh
[2]School of Psychology, University of Southampton

Abstract

Mind-wandering refers to a phenomenon of having thoughts unrelated to the task at hand. Its occurrences have been documented across different activities in both experimentally controlled and real-life situations. Furthermore, evidence suggests that one's mind may start to wander at will and in moments when we preferred it did not. Indeed, mind-wandering has been linked to deterioration of performance in a number of activities. One example is compromised text comprehension: Experiencing mind-wandering episodes is unconducive to efficient reading.

One way we could hope to attenuate the negative influence of mind-wandering on performance is to recognize it and avoid it. However, because one may not be aware that their mind has wandered, we need to rely on external means of discovering mind-wandering. Unfortunately, current state-of-the-art methods of detecting mind-wandering are imprecise, impractical, and obtrusive. In this dissertation, I attempt to ameliorate some of these methodological deficiencies by developing an alternative, a completely unobtrusive way of detecting mindless reading (i.e., mind-wandering during reading).

The ability to read is the sine qua non of daily life in literate societies and has been studied for over 40 years. However, mindless reading literature is far less voluminous. Because of that, I approach mindless reading detection by first systematically studying mindless reading itself thus expanding our understanding of this still nebulous cognitive phenomenon.

Eye movements play a central role in this work. Interestingly, even though text comprehension may cease entirely during mindless reading, eyes of mindless readers move remarkably similar to those of mindful readers. Despite that, my results suggest that eye movements can be used to successfully disentangle these two modes of reading.

Contents

Preface

It would be impossible for one person alone to have accomplished a dissertation like this one. Fortunately, during my PhD studies I have met many fine people who helped me along the way and I would like to take this opportunity to recognize several of them.

I would like to thank my advisors, Marek Druzdzel and Erik Reichle, for always being available and supportive, for being a well of knowledge, and for being good friends. I am very grateful to Marek for letting me stay in his research lab where I spent most of my PhD time. That lab will always have a special place in my heart, but not because of the comfortable couch but rather because of the lab mates I met during my tenure there. Of all of them, I have been the closest friends with Mark Voortman and Martijn de Jongh. The three of us played many a good game of StarCraft 2 and I really appreciate their valuable comments on my work and help with the technical aspect of it even though not all of that has ended up being a part of this document. I would also like to thank Peter Brusilovsky for his help and collaboration. Last but not least, I would like to extend my very special thanks to Rosta Farzan for her fellowship and support.

By deciding to pursue a PhD I had to part with many friends and family and acknowledge that I would no longer be able to see them as often as I would like, a dear price to pay if there is one. Keeping in touch has not always been easy either and because of that I would like to thank Malgorzata Kirszke for being such a good sister and Adrian Aluk-Rowinski for being a great friend in spite of the distance and time difference. Finally, I would like to give a special recognition to my parents, Regina and Leszek Loboda, for their attention and love.

Chapter 1

Introduction

Reading is an activity central to learning and non-spoken communication, both of which have become an integral part of the personal and professional life in developed societies. As such, it has attracted a great deal of research interest. Studying reading, though, is inherently challenging because reading requires mastering several perceptual, cognitive, and motor skills, and demands a high degree of coordination among them. In fact, even though the earliest records of writing go back some 5000 years (Rayner & Pollatsek, 1989; Robinson, 1995), it has been argued that learning to read involves becoming adept at the most difficult skill for which the human brain is not biologically programmed (Huey, 1908; Rayner & Pollatsek, 1989; Reichle et al., 1998). This, however, also makes reading a natural choice of a task to study how cognition (e.g., word identification), perception (e.g., visual encoding), and motor control (e.g. programming of eye movements) interact. Furthermore, reading has well defined task demands which make it especially amenable to experimental control and analysis.

Mind-wandering refers to a phenomenon of having thoughts unrelated to the immediate task being performed (Antrobus, 1968; Giambra, 1995; Singer, 1966; Smallwood & Schooler, 2006; Teasdale et al., 1995; Wegner, 1997). Its occurrences have been documented across different tasks in both experimentally controlled and real-life situations (Kane et al., 2007). While it may serve important functions, it has been linked to deterioration of performance in a variety of tasks. For example, mind-wandering has been reported to compromise text comprehension (Sayette, Reichle, & Schooler, 2009; Sayette, Schooler, & Reichle, 2009; Schooler, Reichle, & Halpern , 2004; Smallwood et al., 2008; Schooler, McSpadden, & Reichle, 2009). Thus, if we could correctly identify episodes of mind-wandering and attenuate their negative impact on performance, we might hope to improve overall reading comprehension.

Unfortunately, the current state-of-the-art method of detecting mindless reading, which combines *thought sampling* (see Smallwood & Schooler, 2006) and self-monitoring instructions, is quite limited. In that method, individuals performing a task are randomly probed to check whether they were mind-wandering or not (which allows for the discovery of the so-called *probe-caught mind-wandering*). Additionally, they are instructed to report when they notice their mind-wandering themselves (which allows for the discovery of the so-called *self-caught mind-wandering*). Because people may not be aware of being engaged in mind-wandering this method is inherently inaccurate in that it may either over- or underestimate the incidence of mind-wandering. Moreover, the instruction to self-monitor may itself affect the frequency of attention lapses. Besides, repeatedly interrupting people to ask if their attention is on- or off-task in too impractical for most real-life situations (be it reading or driving an automobile).

In the current research, I avoid the aforementioned methodological shortcomings by developing an unobtrusive method for detecting *mindless reading* (i.e., mind-wandering in readers). In pursuit of this goal, I first study mindless reading itself thus adding to the still small but expanding literature on that subject. To that end, I investigate which factors (e.g., individual differences) influence a person's propensity to lapse into mindless reading. I also evaluate the relationship between mindless reading and text comprehension. Lastly, I investigate the differences in eye movements of normal and mindless readers. More specifically, I study the prevalence of off-screen fixations and extreme fixation-durations during both normal and mindless reading. Additionally, I weigh the evidence in support of lag, immediacy, and successor effects during both of these "kinds" of reading. This investigation is important in that it informs my subsequent attempts to develop statistical models which use eye movements to identify the moments of a reader's inattention.

Using eye movements to understand and explain how the meaning of text is processed cognitively is extremely informative because it imposes no cognitive constraints (e.g., instructions to verbalize one's thoughts may introduce self-monitoring) other than those of the task itself. Furthermore, the measurement of eye movements yields a high-resolution time-series data that have been proven to

be sensitive to a plethora of behavioral and cognitive factors. In short, the method has proven to be an extremely useful way of studying reading and has been used for that purpose for over 30 years (for a comprehensive review, see Rayner, 1998). Consequently, eye movements are the form of bio-measurement ideally suited for unobtrusive detection of interruptions to the already well-understood process of reading.

Due to the ubiquity of mind-wandering and the importance of reading in literate societies, the proposed research is expected to have important practical ramifications for any application area that involves reading. For example, although not everyone is fond of reading text displayed on a computer screen, there certainly are situations when most of us would be inclined to do so in order to learn something more efficiently. A student studying for an exam under time pressure is a good example that illustrates how automated instruments for detecting mindless reading might enhance learning and provide one method of improving general education.

Another example involves e-learning environments which are devoid of the student-instructor type of interactions that are normally available in traditional learning environments where teachers ordinarily repeat parts of the material or suggest re-reading upon noticing a student's lack of attention. Developing automated methods of detecting mindless reading and alerting students to their attention lapses could help to imitate this aspect of a real learning environment. It might also provide the means for the computer to let an instructor know of the possible comprehension difficulties a student might be having, or even take autonomous actions to address such difficulties.

One final practical example concerns *user modeling* which attempts to infer unobservable information about a user from observable information about them (Zukerman & Albrecht, 2001). Detecting mind-wandering in readers by looking at their eye movements fits this description very well. Because mind-wandering has been associated with deterioration of text comprehension, it is a behavior that could be naturally captured by a *student model* (i.e., a representation of a student's abilities, skills, knowledge, and educational goals). A student model empowered with that new information could provide more accurate predictions for subsequent personalized instruction for a user of an intelligent tutoring system.

This work is of theoretical significance to a number of issues surrounding reading. For example, the application of my work will inform further development of models of eye-movement control during reading (Reichle, Rayner, & Pollatsek, 2003). Moreover, detecting mindless reading may have implications for the interpretation of scientific findings related to reading. That is, if subjects are instructed to read larger passages of text, how will an investigator ensure that mindless reading will not affect the results? In such more ecologically valid settings, detecting episodes of mind-wandering may be necessary to ensure the validity of experimental findings. Additionally, because reading provides a fertile ground for studying the interaction between external stimuli and internal processes of the mind, the study of reading allows one to gain a deeper insight into the general nature of the so called "eye-mind link" (Reichle, 2006). As a result, this work might help to shed light on the more general nature of visual cognition (Findlay & Gilchrist, 2003). Finally, identifying the instances when attention is decoupled from the reading task may help in understanding the architecture of the mind and elucidate the nature of constructs as elusive as consciousness and meta-consciousness (Schooler, 2002).

Of course, this research may also inform the development of an instrument for detecting mind-wandering in tasks other than reading. It is not difficult to imagine situations in which human errors may be costly. Air traffic control, nuclear power plant monitoring, or driving are all good examples of tasks requiring sustained levels of vigilance where errors can have seismic ramifications.

The reminder of this dissertation is structured as follows. In Chapter 2, I briefly summarize what has been documented so far about eye movement behavior during reading (Section 2.1), mind-wandering in general (Section 2.2), and mindless reading in particular (Section 2.3). It is there that I also talk about one of the most important individual difference, working memory (Section 2.4). In Chapter 3, I discuss data acquisition which includes a detailed description of my experiment. In Chapter 4, I discuss analyses I performed in order to better understand evidence pertaining to the differences between normal and mindless reading present in the data I have collected. Finally, in Chapter 5, I discuss my attempts to detect mindless reading.

Chapter 2

Background

2.1 Eye Movements During Reading

2.1.1 Variables

The most widely used classification of factors affecting eye movements during reading discriminate among variables with respect to how early in processing their influence is manifested. That is, the variables are associated with successive stages of text comprehension: *Pre-lexical*, *lexical*, and *post-lexical*. As I discuss in Section 2.3, some of these effects should presumable not be observed when a reader is mind-wandering.

The effects of these variables are normally observed through a number of different dependent variables. For example, as far as *first-pass reading* is concerned *inspection durations* and *inspection probabilities* are used. Measures related to *regressive movements* (e.g., saccades that move the eyes back to earlier parts of the text) are also often employed. Some authors allow for an *early*, *late*, *sometimes-early-sometimes-late* classification (e.g., Clifton, Staub, & Rayner, 2007). Still others propose a general dichotomization between *global averages* (e.g., mean reading rate in word-per-minute) and *word-based measures* (e.g., first-fixation durations; Reichle, Rayner, & Pollatsek, 2003). A good treatment of these issue is provided by Rayner (1998) and Reichle, Rayner, & Pollatsek (2003). Below, I enumerate and describe the most important word variables, as well as give references to studies which investigated them.

Pre-lexical (sublexical, perceptual)

These low-level variables represent the influence of text at the level of visual stimuli and/or as modulated by certain oculomotor (e.g., preferred saccade length; McConkie et al., 1988) and/or cognitive biases (e.g., *the perceptual span*; Den Buurman & Boersma, 1981; McConkie & Rayner, 1976; Nuthmann, Engbert, & Kliegl, 2005; Rayner, 1979; Rayner, Well, & Pollatsek, 1980; Rayner et al., 1982) and include:

- *Word length* (*word boundary*)
  Length refers to the number of characters in a word, with shorter words attracting fewer, shorter fixations.
  Kliegl, Nuthmann, & Engbert (2006); Morris, Rayner, & Pollatsek (1990); O'Regan (1979, 1980); Rayner (1979); Rayner & Morris (1992)

- *Landing site*
  Landing site refers to the ordinal number of the character within a word that receives the initial fixation made on that word. The space preceding a word is counted towards that word and its ordinal number is zero.
  Kliegl, Nuthmann, & Engbert (2006); Rayner (1998)

Lexical (encoding)

These variables represent the influence of *lexical processing*, i.e., the process of activating the basic orthographic-to-sound/meaning connections (lexical entries) that is necessary to identify printed words (Pylkkanen, 2007) and include:

- *Word (log) frequency*
  Frequency refers to how often a work occurs in printed text, with common words being recipients of fewer, shorter fixations.
  Altarriba et al. (2001); Henderson & Ferreira (1990, 1993); Inhoff & Raynes (1986); Juhasz & Rayner (2006); Just & Carpenter (1980); Kliegl, Nuthmann, & Engbert (2006); Rayner & Raney (1996)

- *Word familiarity*
  Familiar words are processed more rapidly than unfamiliar ones.
  Chaffin, Morris, & Seely (2001); Juhasz & Rayner (2003); Williams & Morris (2004)

- *Word type*
  Text contains content words (nouns, verbs, and adjectives; e.g., "bicycle") and function words (articles, conjunctions, prepositions, and pronouns; e.g., "the"). Typically, about 80% of content words and

about 20% of function words are fixated.
O'Regan (1979, 1980)

- *Number of meanings (lexical ambiguity)*
  Words with many meanings require longer processing than those with fewer or one meaning, but processing times are modulated by prior context. For example, preceding context may render one of the meaning dominant and thereby result in fixation duration shorter than that which would be expected otherwise.
  Duffy, Morris, & Rayner (1988); Folk (1999); Rayner & Duffy (1986); Rayner & Frazier (1989)

- *Age of acquisition*
  The earlier in life the meaning of a words is learned the less processing it requires.
  Juhasz (2005); Juhasz & Rayner (2006)

- *Word morphology*
  How a word is constructed determines processing times. For example, morphemes more informative with respect to the overall meaning of a word are fixated longer.
  Juhasz et al. (2003); Hyönä & Pollatsek (1998)

- *Digits*
  Words which are digits require more processing time.
  Just, Carpenter, & Woolley (1982)

Post-lexical (superlexical, postaccess, linguistic)

These higher-order variables represent post-lexical integrative processes during reading (e.g, the integration of syntactic information across successive fixations) and include:

- *Word (logit) predictability* (*contextual constraints*)
  The context in which a word appears in a sentence determines how predictable it is, with more predictable words being fixated shorter or even skipped.
  Balota, Pollatsek, & Rayner (1985); Binder, Pollatsek, & Rayner (1999); Ehrlich & Rayner (1981); Kliegl, Nuthmann, & Engbert (2006); Rayner & Well (1996); Rayner et al. (2004); Schustack, Ehrlich, & Rayner (1987)

- *Plausibility effects*
  Whether a word is plausible or not in a given sentence affects the processes which integrate it into the sentence context.
  Clifton, Staub, & Rayner (2007)

- *Word position*
  Clause- and sentences-terminal words are associated with longer processing times than other words.
  Just, Carpenter, & Woolley (1982)

2.1.2 Eye Movement Control

There is substantial evidence indicating that the decisions about *where* and *when* to move the eyes are largely independent of one another (Rayner & McConkie, 1976; Rayner & Pollatsek, 1987; Reichle et al., 1998; Underwood, 2005). The three major factors influencing the destination of a progressive saccade are a word's boundaries, its length, and the distance of the prior saccade launch site (Rayner, 1998). Compelling evidence also suggests that the duration of a fixation is influenced largely by linguistic properties of words. The two most important variables that have such effect are a word's frequency of occurrence in printed text and its within-sentence predictability (Reichle, Rayner, & Pollatsek, 2003). The decision to move the eyes off a word is triggered by lexical access (or a preliminary state of lexical processing; e.g., a *familiarity check*; Reichle et al., 1998), but other higher-level (e.g., post-lexical) processes may also intervene when something does not "compute," often resulting in pauses and/or regressions (Rayner, 1998).

2.1.3 Models of Eye Movement Control

The increasing understanding of reading have allowed reading researchers to propose models of eye-movement control. One of the most general classifications of these models discriminates them with respect to their assumption about what processes guide eye movements. For example, *cognitive* (or *processing*) *models* maintain that eye movements are under cognitive control of some form (Just & Carpenter, 1980, 1987; Morrison, 1984; Rayner & Pollatsek, 1989; Salvucci, 2000; Thibadeau, Just, & Carpenter, 1982). In those models, lexical processing (or other ongoing comprehension processes) is assigned a major role in influencing when the eyes move. On the other hand, *oculomotor models* posit that oculomotor factors control eye movements and any influences of lexical access are manifested indirectly, when a reader encounters processing difficulty (O'Regan, 1990, 1992; Reilly & O'Regan, 1998; Suppes, 1990; Supper, 1994; Yang & McConkie, 2001). In their purest form, oculomotor models do not allow for the effects of linguistic processing. These two classes of models form extremes of a continuum of the oculomotor-cognitive dimension (e.g., see Reichle, Rayner, & Pollatsek, 2003; see also the 2006 special issue on *Cognitive Systems Research*).

Another important way of discriminating between models of eye-movement control in reading is based on the assumptions they make about the attention allocation (or shift). *Sequential-attention-shift (SAS) models* postulate the existence of the *attention spotlight* (Posner, 1980) which moves from one word to the next in a strictly serial manner (e.g., the E-Z Reader model; Reichle et al., 1998; Re-

ichle, Rayner, & Pollatsek, 1999, 2003; Reichle, Warren, & McConnell, 2009; Reichle, 2010). These models assume that lexical processing is completed on one word at a time. In contrast, *guidance-by-attentional-gradient (GAG) models* assume that attention is distributed as a gradient and thus accommodate parallel word processing (e.g., the SWIFT model; Engbert, Longtin, & Kliegl, 2002; Engbert et al., 2005; Kliegl & Engbert, 2003).

2.2 Mind Wandering

Mind wandering (Antrobus, 1968; Giambra, 1995; Singer, 1966; Smallwood & Schooler, 2006; Teasdale et al., 1995; Wegner, 1997) (also referred to as *daydreaming*, Singer, 1966; *attentional lapses*, Robertson et al., 1997; and *stimulus-independent thoughts*, *SITs*, Mason et al., 2007) refers to a phenomenon of having thoughts unrelated to the immediate task that is being performed. Interestingly, one may not be aware of their mind-wandering and the awareness, or lack thereof, is central to the distinction between *tuning out* versus *zoning out* (Schooler, Reichle, & Halpern , 2004; Smallwood et al., 2008). That is, tuning out is the experience of having off-task thoughts while being aware of having them. On some level, tuning out may reflect instances of deliberate mind-wandering. In contrast, zoning out refers to episodes of mind-wandering that is typically unintentional and that onsets and continues without being noticed (Schooler, Reichle, & Halpern , 2004). This distinction touches on the notion of *meta-consciousness* or *meta-awareness* (Schooler, 2001, 2002; Starr & Rayner, 2001) which is absent in the case of zoning out.

Although the mind is not attending to the immediate task during mind-wandering, it does seem to be occupied with rich thoughts. For example, Schooler, Reichle, & Halpern (2004) found that, during intervals of *mindless reading* (i.e., mind-wandering during reading), 27% of participants were thinking about school-related topics, 19% about fantasies, 11% about themselves, and 18% about nothing at all. They also reported that they were thinking about material related to what they were reading less than 3% of the time (experiment 1).

2.2.1 Magnitude

All of us experience episodes of attention-lapse – the phenomenon is ubiquitous. Reports from laboratory tasks indicate that about 15-50% of a participant's time is spent mind-wandering, depending upon the task: for example, 15% during tasks that involve fluency and encoding (Smallwood, Obonsawin, & Heim, 2003), 20% during reading (Schooler, Reichle, & Halpern , 2004), and 50% during simple signal detection tasks (Antrobus, 1968; Giambra, 1995; Smallwood, O'Connor, et al., 2004). Those proportions may be different in day-to-day living (Smallwood & Schooler, 2006), although a recent study by Kane et al. (2007) suggest that about 30% of our normal waking life is consumed by mind-wandering.

2.2.2 Promoting Factors

Schacter (2001) argued that preoccupation with distracting issues or concerns may be a cause of mind wandering. Evidence reviewed by Smallwood & Schooler (2006) indicates that the mind tends to wander when the primary task does not require executive control. The instances of mind-wandering are more likely to occur in the context of well-practiced tasks in a variety of contexts (Antrobus, 1968; Cunningham, Scerbo, & Freeman, 2000; Giambra, 1995; Smallwood, Baracaia, et al., 2003; Smallwood, Davies, et al., 2004). Even though time on task tends to increase the frequency of mind-wandering episodes, that effect does not generalize to all types of tasks (Smallwood & Schooler, 2006). For example, number of occurrences of mind-wandering seems not to be affected by fatigue (Teasdale et al., 1995). It is, however, affected by mood. The mind tends to wander more when mood is low; for example it is more likely to happen in the case of depressed individuals (Smallwood et al., 2007). The frequency of stimulus-independent thoughts is higher for perceptual as compared to more conceptual processing (McVay & Kane, 2007). Cigarette craving (Sayette, Schooler, & Reichle, 2009), and alcohol consumption (Sayette, Reichle, & Schooler, 2009) both promotes mind-wandering and reduces the likelihood of noticing these mental lapses.

2.2.3 Functions and Effects

Klinger (1999) suggested that instances of mind-wandering may have implications for creative problems solving. This view is shared by Smallwood & Schooler (2006), who suggested, that when the mind wanders, controlled processing is hijacked in the service of an internally-relevant goal. Klinger (1971) allowed a possibility that task-unrelated mental activity can facilitate performance on mundane tasks by enabling an individual to maintain an optimal level of arousal. Its function would then be similar to the nystagmus of an eye – keeping the system from becoming fatigued. Yet another possibility is that mind-wandering is involved in consolidating one's present, past, and possible future experiences (Tulving, 1985; Maguire, 2001; Cabeza et al., 2004; Vincent et al., 2006).

However, task-unrelated thoughts come at a cost to task performance. During mind-wandering, signal detection is poor (Robertson et al., 1997; Smallwood, Davies, et al., 2004), encoding is superficial (Seibert & Ellis, 1991; Small-

wood, Baracaia, et al., 2003; Smallwood et al., 2007), and reading comprehension is compromised (Schooler, Reichle, & Halpern , 2004).

In summary, thoughts that the wandering mind produces are sometimes useful. Nevertheless, there are many situations in which mind-wandering is maladaptive. Finally, the mind may generate task-unrelated thoughts simply because it evolved the ability to manage mental tasks in parallel (Mason et al., 2007).

2.2.4 Measuring

Several ways of measuring mind-wandering have been proposed. The most widely used one is known as *thought sampling*. Thought sampling is related to the *experience sampling* procedure (Hurlburt, 1993) which attempts to sample the content of participants' experiences in ecologically valid settings. Physiological methods (e.g., using electroencephalograms to detect mind-wandering; Cunningham, Scerbo, & Freeman, 2000) are also available, but they are intrusive and unreliable. Retrospective measures of thought sampling, such as thought listing (Seibert & Ellis, 1991) and questionnaire measures of off-task thinking (Smallwood, O'Connor, et al., 2004) have also been used, but they confound awareness and memory of mind-wandering.

Thought sampling can contribute to the detection of *self-caught* and *probe-caught* mind-wandering. *Self-caught mind-wandering* is detected by first explaining the concept of task-unrelated thoughts to the participant, and then instructing them to signal when they detect they were engaged in it. The biggest drawback of this approach is that some of the mind-wandering events will go unnoticed (Smallwood & Schooler, 2006).

Probe-caught mind-wandering employs questions about having been mind-wandering, or so called *thought probes*. For instance, probes employed by Smallwood et al. (2008) asked participants: "Just prior to being asked, was your attention on- or off-task?" (p. 1146). Results obtained by Antrobus, Singer, & Greenberg (1966), Antrobus (1968), Antrobus et al. (1970), and Giambra (1989) suggest, that probes should be administered every 15-30 seconds in order to avoid crudeness of the measurement on the one end and too much interference with the task on the other end.

The probe-caught method provides estimates of how frequently mind-wandering occurs, while the self-caught method provides information about the degree of self-awareness of mind-wandering (Smallwood & Schooler, 2006). Therefore, depending on the research objective, an investigator may choose to employ either or both methods. Schooler, Reichle, & Halpern (2004) suggest that employing either type of detection method does not affect the likelihood of the mind to wander (experiment 2).

2.3 Mindless Reading

Mindless reading refers to episodes when we move our eyes across text while not attending to it. Schooler, Reichle, & Halpern (2004) defined is as a state during which "your eyes may continue moving across the page, the phonology of the words may continue sounding in your head, yet your mind may be elsewhere" (p. 203). The literature addressing this very problem is far from voluminous. Despite the little attention this problem has received, mindless reading is a ubiquitous phenomenon, and is experienced even by skilled readers (Glebnerg, Wilkinson, & Epstein, 1982).

In a recent chapter, Schooler, Reichle, & Halpern (2004) reviewed prior research relevant to zoning out during reading. They identify two general lines of related work. First, *comprehension monitoring* (e.g., Brown, 1980), also referred to as *meta-comprehension* (e.g., Maki & Berry, 1984) or *self-regulated comprehension* (e.g., Hacker, 1998), which shows that meta-cognitive monitoring strategies have important positive implications for reading performance. Second, *task-unrelated images and thoughts* (e.g., Giambra, 1995; Shaw & Giambra, 1993), which has attempted to shed light on the nature and causes of daydreaming. However, as explained by Schooler, Reichle, & Halpern , the applicability of findings of both of those areas of investigation to reading is not straightforward.

In a study on "mindless reading"[1] which attempted to compare eye movements during reading versus scene perception, Vitu et al. (1995) introduced a paradigm called *z-reading*. In this paradigm, participants viewed "text" in which all of the letters had been replaced with the letter "z", but with punctuation and spacing preserved. Participants were instructed to scan the text as if they were reading, and the observed eye movements showed some similarity to those observed during normal reading. The same paradigm was subsequently employed by Rayner & Fisher (1996) and, more recently, Nuthmann (2005). Rayner & Fisher (1996) expressed interest in a true experimental study of mindless reading (i.e., of the type that is the focus of this proposal), but also reflected on the anticipated difficulty associated with such an undertaking. It is interesting to note, that their results challenged the original conclusions of Vitu et al. (1995); the eye movements observed during z-reading tended to be longer in duration than those observed during normal reading, suggesting that some mechanism other than lexical processing is driving the eyes during z-string reading.

The significance of z-reading in the context of mind-wandering is underlined by a suggestion of Schooler, Re-

[1] I use quotes to denote the disengagement of lexical processing through manipulation of the properties of text (deprivation of most of semantic features; e.g., "reading" strings of the letter *z*), as opposed to not attending to normal text during mind-wandering episodes.

ichle, & Halpern (2004) – that the link between eye movements and lexical processing may be broken during mindless reading. Since z-reading provides information about the oculomotor behavior in the absence of lexical processing demands, there may be a parallel between findings of studies employing that paradigm and zoning out. However, as indicated by Rayner & Fisher (1996), the fact that their subjects (college undergraduates) had many years of reading experience may have influenced their eye movements in a way that made them resemble eye movements indigenous to normal reading.

Schooler, Reichle, & Halpern (2004) grant a possibility that because word identification in highly skilled readers is a largely automatic process (Rayner & Pollatsek, 1989), any processing beyond the lexical level may stop during zoning-out episodes. More specifically, one might predict to observe frequency effects in the absence of predictability effects (or any other effects associated with post-lexical integrative processes).

Mind-wandering has also been reported to compromise text comprehension. Schooler, Reichle, & Halpern (2004) and Smallwood et al. (2008) discovered that the tendency to zone out was associated with a particularly low level of attention to the text. Schooler, McSpadden, & Reichle (2009) found that failure to notice "gibberish" (i.e., text in which the order of content words was randomly shuffled) tended to coincide with being caught mind-wandering. Sayette, Schooler, & Reichle (2009) found that smokers in a state of craving nicotine (because of not being allowed to smoke) tended to perform worse on a text comprehension test than their non-craving counterparts.

In what seems to be the most elaborate and ecologically valid study of mindless reading conducted so far, Reichle, Reineberg, & Schooler (2010) had participants read an entire novel on an eye-tracker using the standard self- and probe-caught measures of mind-wandering. This experiment documented several interesting findings, including the fact that mind-wandering episodes can be quite long in duration (e.g., 1-2 minutes) and that mind-wandering seems to become more profound over time. The individual fixation durations were longer and less modulated by lexical variables (e.g., word frequency) when subjects were zoning out. This suggests that the influence of variables that guide eye movements during normal reading may partially disappear during intervals of mind-wandering. Reichle, Reineberg, & Schooler also observed an increasing number of off-text fixations immediately before (2.5 seconds) subjects reported zoning-out. This behavior may be indicative of subjects gaining meta-awareness of having task-unrelated thoughts.

The association of zoning-out episodes with poor text comprehension also has some implications for out understanding of meta-cognition. For example, if people understand that zoning out is incompatible with successful reading, the fact that their minds lapse into mind-wandering without them being aware of it suggests that, at times, people may be unable to inspect the content of their consciousness (Schooler, Reichle, & Halpern , 2004). This is also why it may be reasonable to assume that the many (or perhaps most) mind-wandering episodes in readers are in fact cases of zoning out. Consistent with this hypothesis is the fact that Schooler, Reichle, & Halpern (2004) found that readers were unaware of having been mindlessly reading on about 67% of mind-wandering responses (experiment 1). They also found their subject were zoning out 13% of the 45-min. reading session. Finally, Sayette, Schooler, & Reichle (2009) similarly report a marked decrease of the ability to notice mind-wandering in craving smokers.

2.4 Working Memory

Working memory (*WM*) is one of the central concepts in cognitive psychology. It is responsible for temporary storage and processing of information (for a review see Kintsch et al., 1999). *Working memory capacity* (*WMC*) is one of the most important individual difference measures and is measured with complex span tasks. WMC is distinct from *short-term memory capacity* (*STMC*), which is measured by simple span tasks, such as modified digit span task (Daily, Lovett, & Reder, 2001). STMC reflect primarily domain-specific storage (e.g., verbal materials); in contrast, WMC reflects domain-general executive attention (Engle et al., 1999).

2.4.1 Working Memory Span Tasks

In a *working memory* (*WM*) *span task* (or *complex span task*), a subject is presented with a sequence of items, one at a time, for subsequent serial recall. To suppress the use of compensatory strategies (e.g., subvocal rehearsal) and to ensure the engagement of the processing component of WM, a secondary task is interleaved with the memory retention task.

There are different versions of WM span tasks. They differ with respect to the nature of their primary and the secondary tasks. The primary tasks that are often used include reading or listening to sentences, solving arithmetic problems, counting objects displayed in different colors, deciding whether or not letters are mirror images or not, and judging whether spatial patterns are symmetrical. Digits, letter, words, shapes, and spatial locations are used as to-be-remembered items (Unsworth et al., 2009). The most popular WM span task are the reading span, the operation span, and the counting span (e.g., Conway et al., 2005).

Performance on WM span tasks correlates with perfor-

mance on lower-lever attention and perception tasks. For example, individuals with scores in the lower quartiles on the latter tasks have difficulty resisting the attention capture of an exogenous cue in the anti-saccade task (Kane et al., 2001; Unsworth, Schrock, & Engle, 2004), have difficulty constraining their attention to discontinuous regions of space (Bleckley et al., 2004), are slower to constrain their focus of attention in a flanker task with incompatible distracters (Heitz & Engle, 2007), make many more errors in a Stroop task (Kane & Engle, 2003), are more vulnerable to proactive interference (Kane & Engle, 2000).

WMC also predicts performance a wide range of higher-order cognitive tasks, such as reading and listening comprehension (Daneman & Carpenter, 1983; Daneman & Merikle, 1996), language comprehension (King & Just, 1991), following oral and spatial directions (Engle, Carullo, & Collins, 1991), vocabulary learning from context (Daneman & Green, 1986), note taking in class (Kiewra & Benton, 1988), writing (Benton et al., 1984), reasoning (Barrouillet, 1996; Kyllonen & Christal, 1990), hypothesis generation (Dougherty & Hunter, 2003), bridge playing (Clarkson-Smith & Hartley, 1990), and complex-task learning (Kyllonen & Stephens, 1990). Individual differences in WMC predict *general fluid intelligence* (Kane et al., 2001) or the ability to think logically and solve new problems which itself is relatively stable across the lifespan (Conway, Kane, & Engle, 2003). WM span scores are also strongly related to standardized test scores, such as SAT (Daneman & Merikle, 1996).

2.4.2 Reading Span Task

In the *reading span task* (*RSPAN*; originally devised by Daneman & Carpenter 1980), subjects are presented with a sequence of to-be-remembered letters. Between each letter, they are also presented with sentences that they are asked to read and then assess whether or not a sentence is correct.

Performance on RSPAN has been found to be a good predictor of word reading times in sentence comprehension (Just & Carpenter, 1992). Due to the similarity between reading and the task of assessing correctness of sentences, RSPAN seems especially well suited for studies involving reading. However, it should be noted that the nature of the secondary task does not seem to matter. For example, Daneman & Merikle (1996) and Turner & Engle (1989) found that operation span predicts performance on sentence comprehension equally well as RSPAN.

2.4.3 Mind Wandering

WMC seems to mediate the relation between mind wandering and cognitive demand (Kane et al., 2007). During challenging tasks, high-span subjects were able to maintain focus and mind-wander less often than did low-span subjects. High-spans mind-wandered more than low-spans when performing easy tasks, suggesting that low-spans could not have any mental resources left to support mind-wandering. Additionally, the ability to perform complex WM tasks is frequently impaired during mind-wandering episodes (Smallwood & Schooler, 2006).

Chapter 3

Data Acquisition and Preparation

3.1 The Experiment

The purpose of the current experiment was to collect behavioral, eye-movement, and mind-wandering data of subjects reading ecologically valid material. 116 native English speaking literate male and female students aged 18 and above were recruited to participate in this study. Subjects were asked to read the first several chapters of Jane Austen's *Sense and Sensibility* novel that was displayed on a computer screen while the subjects had their eye movements recorded by an EyeLink 1000 eye tracker (SR Research, Ltd.). Participants were provided the following definition of zoning out: "At some point during reading, you realize that you have no idea what you just read," and that "not only were you not thinking about the text, you were thinking about something else altogether." Behavior-sampling probes in the form of on-screen questions were presented every 2-4 minutes (time selected randomly from a uniform distribution) to discover whether subjects were mind-wandering without awareness of doing so. Additionally, subjects were asked to promptly self-report mind-wandering episodes by pressing a specified button whenever they caught themselves mind-wandering. Individual differences were collected through a working memory span task and pre- and post-questionnaires. All subjects participated in a text comprehension test. The experiment took about two hours.

The experiment protocol was as follows: Upon arriving at the eye-tracking lab, a subject was administered the reading span task (RSPAN; see Section 2.4.2) task I have implemented[1]. Afterwards, the subject filled out a short questionnaire (see Appendix A) asking about the books they have read (*Sense and Sensibility* being among them) and three of their favorite genres. Next, they proceeded with self-paced reading of the novel, two-to-three chapters at a time, with breaks for forced multiple-choice comprehension questions and eye tracker recalibration. A single page was presented on screen in it's entirety, and each set of chapters consisted of 12-16 pages. The composition of a page was chosen to maximize the size of individual words and the distances between both the words and the lines of text without making pages too short (i.e., presenting too few lines per page). At the end of the experiment, a subject filled out a longer questionnaire (see Appendix B) asking for the following items:

- **Demographics**
  Gender
  Age
  Verbal SAT score

- **Text**
  Sense and Sensibility movie
  Romance a favorite genre
  Drama a favorite genre
  Interestingness of the text
  Focus on text

- **Context**
  Stress
  Fatigue
  Preoccupation (with school, work, and personal problems)
  Craving (what and how strongly)

Answers to questions requiring an indication of magnitude were given on Likert-like scales. The instructions for the RSPAN and reading tasks can be found in Appendices C and D.

3.1.1 Reading Span Task

Subjects went through three trials of set size two, three, four, five, six, and seven, which yielded a total of 18 trials. The order of trials was randomized for each subject. The to-be-remembered item (a single character) was presented in the center of the screen for 1000 ms. Apart from storing the item for later recall, subjects also needed to assess the correctness of sentences, each being 10-15 words in length. Sentence presentation order was randomized for each subject.

[1] http://www.pitt.edu/~tol7/res/research/psych-tests/rspan

Subjects went through an item-only test trial of set size three and a sentence-only test trial of size 11. During the sentence-only test trial average sentence reading time was established for each subject (the very first sentence was excluded from that measurement). That average time (plus 2 standard deviations) was used to time sentence presentation during actual trials. For example, if no answer was provided to the sentence assessment question during that time, that sentence assessment was counted as an error. Subjects were asked to maintain their accuracy on the secondary task at or above 85%. Their current accuracy was shown to them after every trial.

3.2 Data Preparation

The experiment produced large quantities of data. In this section, I describe processing I subjected that data to before engaging in the statistical inference and classification tasks. All automation was done with the Pegasus software (Loboda, 2009).

3.2.1 Regions of Interest

In an eye tracking experiment, it is necessary to define a series of *regions of interest* (*ROIs*; e.g., a word) that are then treated as units of analysis. A unit of theoretical importance in the current experiment was a single word. The eye tracking software paired with the eye-tracker I used was able to transform samples of gaze location[2] into a sequence of fixations, saccades, and blinks. It did not, however, provide support for defining ROIs automatically and matching them to eye-movements and therefore I had to do that myself.

The algorithms I developed worked as follows. First, one ROI per word was defined for each page of text. Additionally to finding coordinates of bounding boxes of all ROIs, my algorithms recognized sentence and clause boundaries and marked the respective ROIs accordingly. Next, to use a visual metaphor, the algorithms overlayed fixations over the ROIs, and for each fixation they found the ROI that fixation was inside of. Because the ROIs did not overlap, it was not possible to have more than one ROI per fixation. It was, however, possible to have no ROIs per fixation (in the case of off-text and off-screen fixations). Note, that the algorithms I implemented provided visual feedback on the definition of ROIs (Figure 3.1) and I went through all the stimuli to ensure all ROIs were defined correctly. Figure 3.2a shows fixations made by one of the subjects over a sample page of text.

[2]The EyeLink 1000 eye tracker samples gaze location 1000 times per second.

3.2.2 Fixations Adjustment

Eye movement recordings from the current experiment were subject to both global and local errors. An example of a global recording error was having all fixations shifted vertically by the same distance. An example of a local recording error is a tendency of fixations made on a single line of text to ascend from the left side of the screen to its center and then descend from the center to its right side forming an arc instead of a straight horizontal line. In an attempt to mitigate the influence of those errors, I performed two types of automatic adjustments. Global adjustments attempted to position the "fixation cloud" so that it best matches the underlying text. This adjustment affected all fixations made on a given stimulus equally (geometrically speaking, all fixations were translated by a vector). Local adjustments attempted to identify fixations that belong to the same line of text taking the sequentiality of fixations into account. Once a sequence of fixations was identified (subject to parameters such as minimum fixation duration, minimum number of fixations, maximum vertical distance between two adjacent fixations, etc.), the local adjustment algorithm made all of the fixations belonging to a given line parallel to the line of text.

3.3 Subject Exclusion

I only excluded four subjects with clear problems in behavioral data. Two subjects bailed out of the study after staying only for about half an hour, enough to get the credit for participation. One other subject reported a staggering number of mindless reading episodes (257; max. for the other subjects was 54, Table 4.2) apparently not understanding the experiment instructions while yet another subject did not report nor was caught mindlessly reading even once. Interestingly, the two subjects who bailed out were not caught mindlessly reading even once and only self-reported it. To sum up, despite the fact that my experiment had the size of 116, 112 is the number of subjects I retained for all my analyses. Note also that missing values in some subject-level variables (e.g., the SAT score; Table 4.2) resulted in lowering that number in some of the inferential models I fitted, but I report the exact N for each model separately.

I visually examined all eye-movements recorded during my experiment and found cases of trials (a single trial consisted of reading of 2-4 chapters, depending on length) with "bad eye-movements" (Figure 3.2b). I also noticed that such weird patterns coincided with the eye-tracker calibration–validation results being marked below "good–good" (with the possible values for both calibration and validation being: "poor," "ok," and "good"). To avoid having to deal with even slightly questionable eye-movement data, in all analyses involving eye-movement

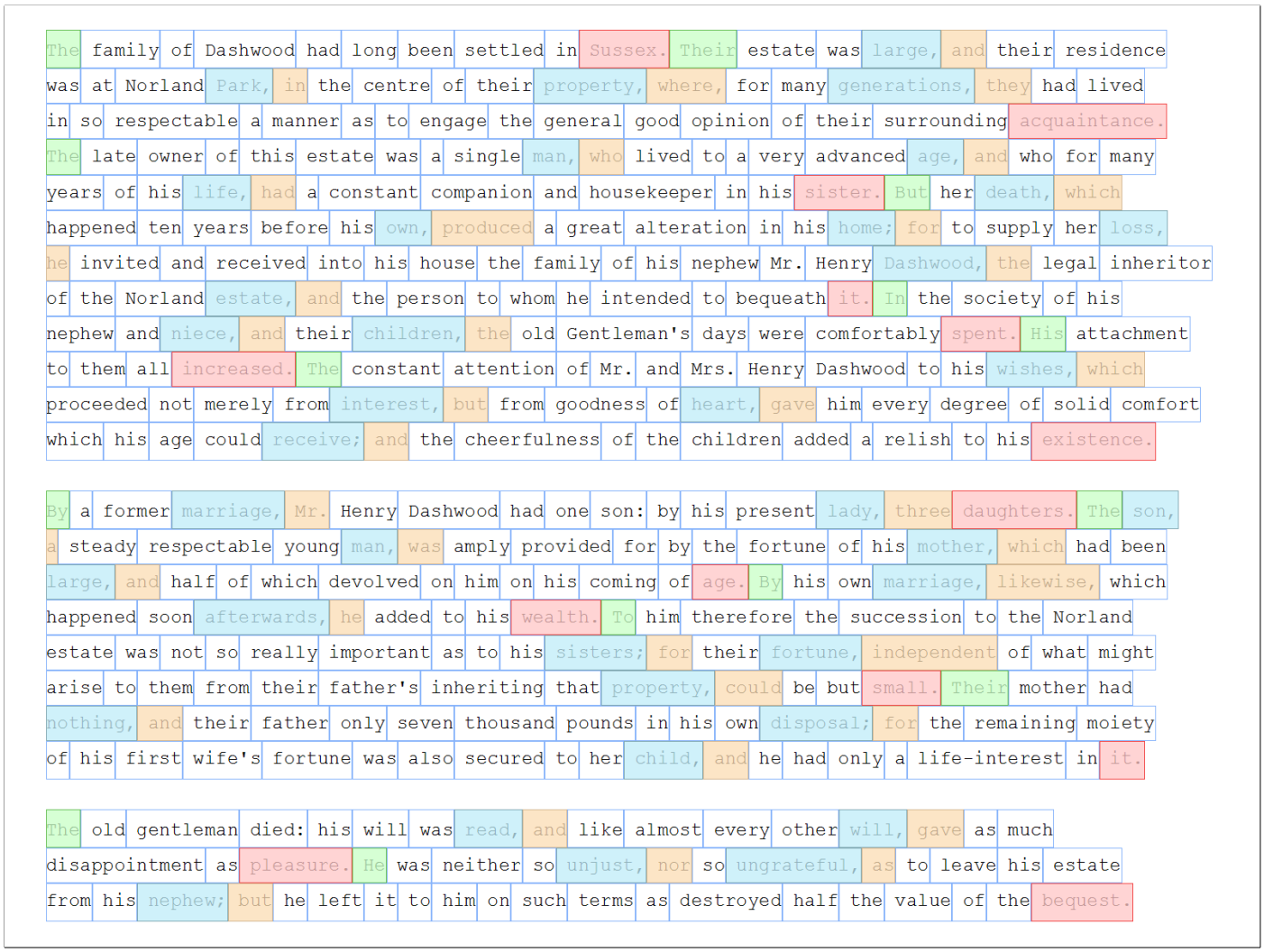
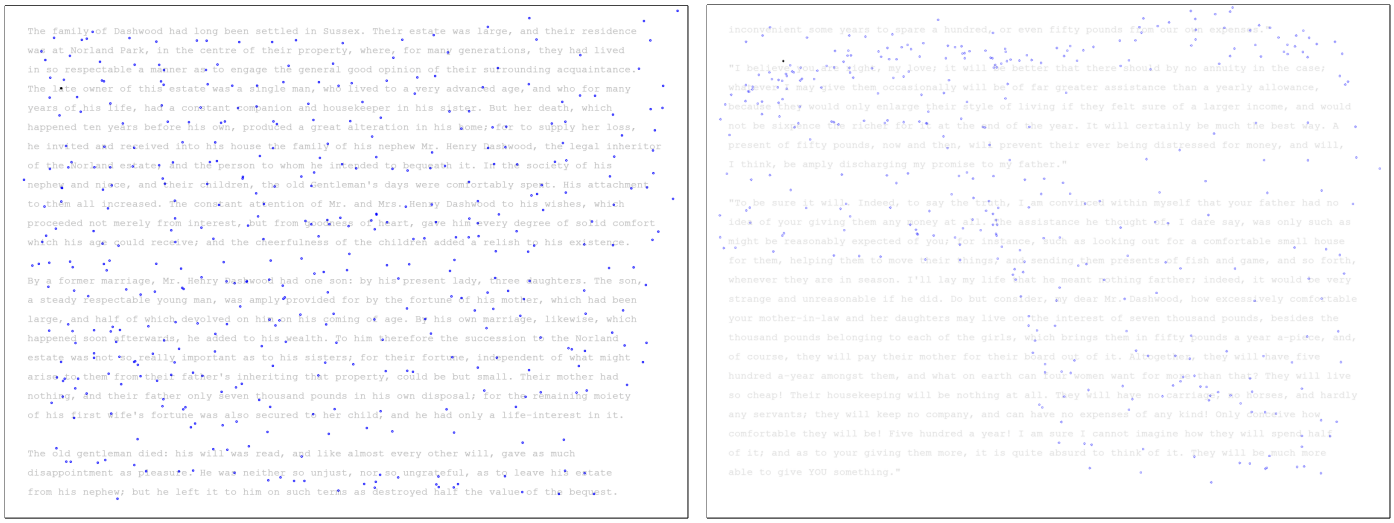


Figure 3.1: A sample stimulus (*Sense and Sensibility*, Chapter 1, Page 1) with regions of interests defined and sentence and clause start and end words highlighted.

variables (including modeling) I used only data from trials with calibration–validation results marked as "good–good." This resulted in exclusion of a total of 14 subjects. There was a problem with one of the experiment computers which resulted in a loss of one additional subject's eye-movement data which for whatever reason did not get recorded (behavioral data for that subject were collected).



(a) Good

(b) Bad

Figure 3.2: An example of good and bad eye-movements. Calibration–validation results were “good-good“ on the left and “good–poor” on the right).

Chapter 4

Investigating Mindless Reading

In this chapter, I discuss on my investigation of relationships between subject variables (e.g., working memory index), word variables (e.g., word length and frequency), eye movement variables (e.g., first fixation duration), and the reading state (i.e., mindful vs. mindless). When carrying out my analyses, my main goal was to understand significant relationships between these variables in order to ultimately use them to decide which variables to include in my classification models (Chapter 5). A few important clarifying remarks follow.

In this research, I deal with three types of events: Catching a subject reading normally (or mindfully), catching a subject reading mindlessly, and having a subject catch themselves reading mindlessly. For brevity, throughout this document I may refer to these three types of events as **N**, **P**, and **S**, respectively. Furthermore, whenever I need to combine both types of mindless reading into what I call "total mindless reading" I may refer to that combination as **PS** or **P+S** events.

When talking about probe-caught and self-caught mindless reading I may refer to them as the two "kinds" of mindless reading. I do that purely for brevity and do not imply any fundamental differences between them. While difference on both the level of temporal patterns of attention decoupling (e.g., how intermittent it is) and in terms of strength of that decoupling most likely exist, these two still are manifestations of the same cognitive phenomenon.

Because the duration of a typical mindless reading episode has not yet been convincingly established, I conduct all analyses involving eye-movement variables assuming different durations of those episodes. To that effect, following Reichle, Reineberg, & Schooler (2010), I use time windows of different sizes to sample eye movements I use in my analyses. If t_0 denotes the onset of a thought-sampling probe or a self-report of mindless reading, then each of time windows I use begins a designated amount of time prior to t_0 and ends at t_0 (i.e., all of them go back in time from t_0). I do not use time windows narrower than one second or wider than 120 seconds.

Note also that because probes were temporarily random, it was not possible for the subjects to predict when they would appear. Because of that fact and because I do not sample any eye movements after the onset of a probe, eye movements that are associated with anticipating a probe or responding to it are unlikely to affect my results. However, the presence of probe-initiated interruptions and the instruction to self-monitor the contents of one's consciousness are two sources of unavoidable contamination in this experiment.

I conducted all analyses I report in this chapter in the Statistical Analysis System versions 9.2 and 9.3 (SAS Institute Inc., 2008). I use both generalized linear and generalized linear mixed models (GLMs, GLMMs, respectively; Nelder & Wedderburn, 1972; McCullagh & Nelder, 1989). Unless indicated otherwise, I report unadjusted type 3 p-values.

4.1 Working Memory

My experiment started with the subjects participating in a complex WM task and that is also where I start presenting the results. There are four schemes of combining the results of a complex WM task into a composite score (Conway et al., 2005): Partial-unit (PU), partial-load (PL), full-unit (FU), and full-load (FL). The partial scheme assigns a fractional score to trials recalled partially while the full scheme assigns 1 for trials recalled perfectly and zero otherwise. Furthermore, the unit scheme sees all trials as equal irrespective of the item size while the load scheme assigns higher score to trials with larger item sizes. The correlations between the four composite scores and the normalized self-reported verbal SAT scores ($id.sat$; four subjects reported the ACT score) are shown in Figure 4.1 and in Table 4.1. I used the following normalization formulae for the SAT and ACT scores

$$id.sat = (SAT - 200)/600,$$
$$id.sat = (ACT - 2)/35.$$

Measures of working memory capacity tend to correlate well with SAT scores Daneman & Merikle (1996). That

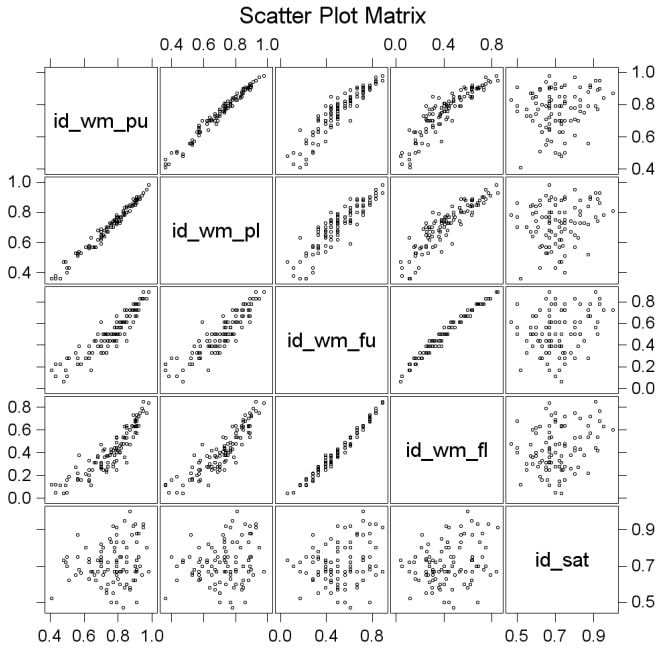


Figure 4.1: Correlations between the individual differences variables (id).

Table 4.1: Correlations between the individual differences variables (id).

	wm.pu	wm.pl	wm.fu	wm.fl	sat
wm.pu	1				
wm.pl	.99	1			
wm.fu	.92	.90	1		
wm.fl	.90	.90	.99	1	
sat	.18	.20	.22	.24	1

was however not the case in the current experiment. Because I did not ask the subjects to bring their SAT scores with them, this low correlation could be a result of poor recollection on their part. In that respect, working memory capacity should be more representative of their mental capacity especially on the day of experiment.

As shown in Table 4.2, the accuracy on the secondary task ($id.wm.acc$) was high ($\mu = .88$, $\rho = .08$). As anticipated from the way in which the four composite scores are calculated, PU has the highest mean, followed by PL, FU, and FL. In the reminder of this dissertation, I use the FL score because it has the highest variance of all four and therefore provides the strongest separation between subjects.

4.2 Subject-Level Variables

Table 4.2 and Figure 4.2 show descriptive statistics and histograms of all subject-level variables which were collected in the experiment (e.g., age) or calculated afterwards (e.g., craving, or $ctx.crv$, discretized into $ctx.crv.3s$). I grouped these variables into the following six groups: demographics (dem), text (txt), reading ($read$), individual differences (id), context (ctx), and eye-movement (em).

4.2.1 Demographics and Text Variables

41% of subjects were males and the age of all subjects was 18–28 with most subjects being 18 or 19 years old. Because the variation in age was very small, I did not include that variable in any of the models. Subjects read, on average, almost ten chapters, or 50 pages of text. Because only 8% of subjects read the book before and only 3% saw the movie, I did not include these two variables in any of the models. Text comprehension was high (68%). Romance was among the top three favorite genres of 26% of subjects, and drama for 12%. The average self-reported interestingness of text was 3.20 (1–7) and self-reported focus was 4.09 (1–7).

Text comprehension in my experiment was lower than that reported by the majority of sentence-reading studies. That is to be expected because of two reasons. First, the amount of material presented to subjects in a natural reading study is much higher then in a sentence-reading study (here 50 pages of text versus typically 50-150 sentences). Second, mindless reading could have contributed to the degradation of text comprehension by taking the attention away from the text which seems less likely to happen in the context of reading individual and typically unrelated sentences. I investigate this claim in more detail in Section 4.6.

4.2.2 Reading Variables

Subjects spent, on average, 72 minutes reading progressing through the text at an average rate of 248 words per minute. During that time, an average subject was probed for mind-wandering about 16 times. About 13 of these probes (47% of all MW events) appeared when the subject was reading mindfully and three of them (12% of all MW events) when they were reading mindlessly. An average subject self-reported about 15 episodes of MW (41% of all MW events).

I calculated the reading speed of each subject by averaging the quotient of the amount of words per page and the total amount of time spent on that page.

Because the proportion of probe-caught MW episodes was small (24%), analyses involving those episodes may

be unreliable, especially the EM analyses. That is why, throughout this dissertation, I treat the results related to the P events with less credibility than those related to the N and S events.

4.2.3 Individual Differences Variables

The SAT score ($id.sat$) was missing for 12 subjects. Therefore, all statistical models with this variable included have the sample size reduced by 12. Working memory capacity is the other individual difference which I have described earlier, in Section 4.1.

4.2.4 Context Variables

Most subjects competed the experiment between 2PM and 6PM (i.e. in the afternoon). The average self-reported stress level was 3.42 (1–7), fatigue was 4.18 (1–7), preoccupation was 9.61 (1–21), and the likelihood of future participation was 3.98 (1–7). Subjects did not crave a lot, on average; 4.22 (0–14). In fact, the histogram of $ctx.crv$ (Figure 4.2) revealed, that about 40% of subjects did not crave at all. To account for that, I created a three-level categorical variable $ctx.crv.3s$: 0–no craving, 1–low craving, and 2–high craving (median split on $ctx.crv > 0$). Two observers performed the experiment, as indicated by the binary variable $ctx.obs$. I check for the observer effect in Section 4.3.

4.2.5 Eye Movement Variables

I properly transition to analyses based on eye-movements in Section 4.9. Here, I only report the total number of blinks made by subjects during reading (thus a variable naturally aggregated at the subject level). The reason I am interested in this variable is because it has been reported that blinking rate was elevated during mindless reading (Smilek, Carriere, & Cheyne, 2010) and I intend to check if the total number of blinks is related to mindless reading.

4.2.6 Multicollinearity

To identify pairs of variables which could cause problems due to multicollinearity, I checked correlations between all variables I was considering using (Table 4.3). Three correlations were larger than or equal to 0.41. As a consequence, I excluded two variables. First, I excluded sex ($dem.sex$) because it was correlated with a more general variable, interestingness of text ($txt.int$; $\rho_{dem.sex,txt.int} = -.41$, $\rho^2 = .17$, $p < .0001$). Second, I excluded stress ($ctx.str$) because it was correlated with fatigue ($ctx.fat$; $\rho_{ctx.str,ctx.fat} = .45$, $\rho^2 = .20$, $p < .0001$) and preoccupation ($ctx.pre$; $\rho_{ctx.str,ctx.pre} = .49$, $\rho^2 = .24$, $p < .0001$). Extra motivation for excluding stress instead of preoccupation was provided by (Schacter, 2001) who found preoccupation to increase the likelihood of mind-wandering and I was interested in checking if that was also the case for my natural reading task.

Despite collecting focus on text ($txt.foc$), I have eventually decided to exclude it because lack of focus and mind-wandering can be used interchangeably. That is, not being focused on text automatically implies that attention has been dissociated from the immediate task context.

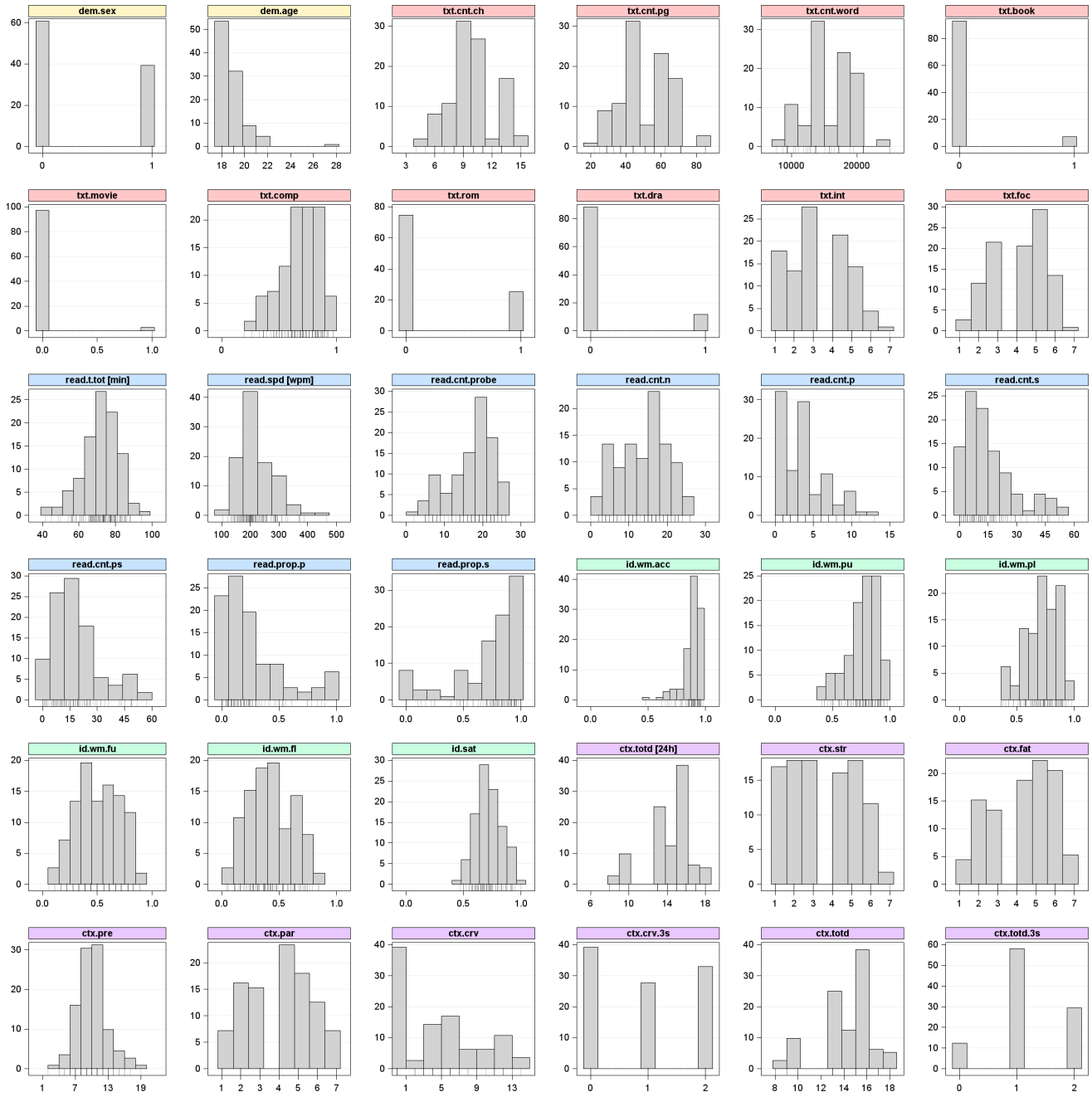


Figure 4.2: Histograms of subject-level variables (apart from $ctx.obs$). Unit or type of the variable are given for some variables in square brackets ([c]ount, [r]atio). The colors of the labels are only a visual cue for grouping variables into categories.

Table 4.2: Subject-level variables.

Variable	N	μ	σ	Min	Max	Description
Demographics						
dem.sex	112	0.39	0.49	0	1	Gender (1–male)
dem.age	112	18.73	1.23	18	28	Age
Text						
txt.cnt.ch	112	9.93	2.28	4	15	Number of chapters inspected
txt.cnt.pg	112	50.55	14.10	22	85	Number of pages inspected
txt.cnt.roi	112	15607.59	3791.92	7749	24989	Number of words inspected
txt.book	112	0.07	0.26	0	1	Read *Sense and Sensibility* book? (0/1)
txt.movie	112	0.03	0.16	0	1	Seen *Sense and Sensibility* movie? (0/1)
txt.rom[1]	111	0.25	0.43	0	1	*Romance* among the top three favorite genres? (0/1)
txt.dra[1]	111	0.11	0.32	0	1	*Drama* among the top three favorite genres? (0/1)
txt.comp	112	0.68	0.17	0.20	0.97	Text comprehension
txt.int	112	3.18	1.47	1	7	Interestingness of text (self-reported; 1-7)
txt.foc	112	4.06	1.36	1	7	Focus on text (self-reported; 1-7)
Reading						
read.t.tot	112	72m 0s	9m 41s	41m 36s	94m 9s	Time spent reading [min] (total)
read.spd	112	220.93	62.08	103.73	466.64	Reading speed [words per min.] ($txt.cnt.roi/read.t.m$)
read.cnt.probe	112	13.25	6.43	2	26	Number of probes
read.cnt.n	112	12.99	6.55	0	26	Number of N events
read.cnt.p	112	3.40	2.85	0	13	Number of P events
read.cnt.s	112	14.62	13.01	0	54	Number of S events
read.cnt.ps	112	18.01	13.10	0	57	Number of MW episodes ($read.cnt.p + read.cnt.s$)
read.prop.pc	112	0.24	0.21	0	1	Probe-caught ratio ($read.cnt.p/read.cnt.probe$)
read.prop.p	112	0.26	0.27	0	1	Proportion of P events ($read.cnt.p/read.cnt.ps$)
read.prop.s	112	0.71	0.29	0	1	Proportion of S events ($read.cnt.s/read.cnt.ps$)
Individual differences						
id.wm.acc	112	0.88	0.08	0.48	0.98	Accuracy on the secondary task
id.wm.pu	112	0.76	0.13	0.41	0.98	Composite score: partial-unit
id.wm.pl	112	0.72	0.14	0.36	0.98	Composite score: partial-load
id.wm.fu	112	0.51	0.19	0.06	0.89	Composite score: full-unit
id.wm.fl	112	0.42	0.20	0.04	0.85	Composite score: full-load
id.sat	100	0.71	0.11	0.47	1	Normalized verbal SAT[1)] score (recalled on the spot)
Context						
ctx.str	112	3.41	1.71	1	7	Stress level (self-reported; 1-7)
ctx.fat	112	4.22	1.62	1	7	Fatigue (self-reported; 1-7)
ctx.pre	112	9.54	2.80	2	19	Preoccupation (self-reported; 0-21)
ctx.crv	112	4.17	4.26	0	14	Craving (self-reported; 0-14)
ctx.crv.3s	112	0.94	0.85	0	2	Craving discretized (0–no craving, 1–low, 2–high)
ctx.totd	112	14.13	2.32	8	18	Time of the day (TOTD) the experiment started at (24h)
ctx.totd.3s	112	0.99	0.49	0	2	TOTD discretized (0–before 12PM, 2–4PM or after, 1–else)
ctx.par	111	3.96	1.68	1	7	Likelihood of future participation (self-reported; 1-7)
ctx.obs	112	0.45	0.50	0	1	The observer (0/1)
Eye movements[2)]						
em.nb.tot	99	1058.21	702.29	99	3271	Number of blinks (total)

[1)] Several subjects reported ACT scores and that was the reason to normalize both SAT and ACT onto the range [0..1]; 12 subjects did not provide any score

[2)] Data based on the subset of subjects that were not excluded due to abnormal eye-movements (see Section 3.3)

Table 4.3: Correlations between non-binary subject-level variables which I considered including (not as an offset) in at least one statistical model. The highlighted high correlations ($> .4$) may indicate multicollinearity problems.

	dem.sex[1)]	txt.rom	txt.dra	txt.int	read.spd	id.wm.pl	id.sat	ctx.str[1)]	ctx.fat	ctx.pre	ctx.crv.3s	ctx.totd.3s
dem.sex[1)]	1											
txt.rom	−.34	1										
txt.dra	−.07	.05	1									
txt.int	**−.41**	.15	.14									
read.spd	−.15	−.05	−.02	−.04	1							
id.wm.pl	−.14	.08	−.08	.09	.06	1						
id.sat	−.16	−.08	.02	.34	.16	.20	1					
ctx.str[1)]	−.11	.00	.04	.01	.01	−.12	−.02	1				
ctx.fat	−.11	.10	−.10	−.08	.05	−.04	−.04	**.45**	1			
ctx.pre	−.12	.07	.06	−.07	.07	.04	−.13	**.49**	.24	1		
ctx.crv.3s	.05	−.03	−.14	−.21	.10	−.18	−.06	.03	.21	.18	1	
ctx.totd.3s	.05	.22	.06	.00	−.12	.13	.04	−.05	−.24	−.04	−.22	1

[1)] Variable excluded due to a high correlation with at least one other variable; see main text for explanation

4.3 Observer Effect

Motivation

The current experiment was run by two experimenters, a male (myself) and a female (my assistant, Adrienne DiFonso). Rosenthal (1998, 2002, 2003) has shown that an observer (or experimenter) can influence the outcome of an experiment through the so called *interpersonal expectancy effects*. These effects are manifested by the observer communicating to a subject, usually unintentionally and nonverbally (e.g., through putting emphasis or showing agitation) the most desirable experiment outcome and thereby affecting their behavior and performance. The fact that the subjects participating in the current experiment were not presented with any clear measure of performance that the experimenter could be excited about makes it less likely for the observer effect to have occurred but that is not enough to dismiss that possibility. Moreover, Singer (1988) reported an increased incidence of TUITs when the sex of the experimenter and subject differed. Investigating these two issues is, in my opinion, worthwhile.

As indicated in Table 4.2, each of the observers handled roughly half of the subjects ($ctx.exp$). To check whether the observer effect is present in the data collected I use two response variables: The total number of mindless reading episodes ($read.cnt.ps$) and text comprehension ($txt.comp$). Contingent upon the results, I will opt to include or exclude the observer variable ($ctx.obs$) from all analyses reported in the remainder of this dissertation.

Models and Data

To investigate the observer effect on the propensity to lapse into mindless reading, I fitted the following model

$$\begin{aligned} read.cnt.ps_i &\sim \text{Poisson}(\lambda_i), \\ \log \lambda_i &= \beta_0 + \beta_1(ctx.obs_{[1]i}) + \beta_2(dem.sex_{[1]i}) \\ &\quad + \beta_3(dem.sex_{[1]i} \times ctx.obs_{[1]i}) \\ &\quad + \log(read.t.tot_i) \\ \text{Var}[read.cnt.ps_i] &= \phi\lambda_i, \\ i &\in \{1, 2, \ldots, 112\}, \end{aligned} \tag{4.1}$$

where i is the subject index. I used time spent reading ($read.t.tot$) as an offset variable to account for differences in exposition times and included a multiplicative scale parameter ϕ on the variance function which mitigated a substantial overdispersion (before: $\sigma^2_{pr} = 3.16$; after: $\sigma^2_{pr} = 1.00$). The residuals panel (Figure 4.3) shows a good fit of the model.

To investigate the observer effect on text comprehension, I fitted the following model

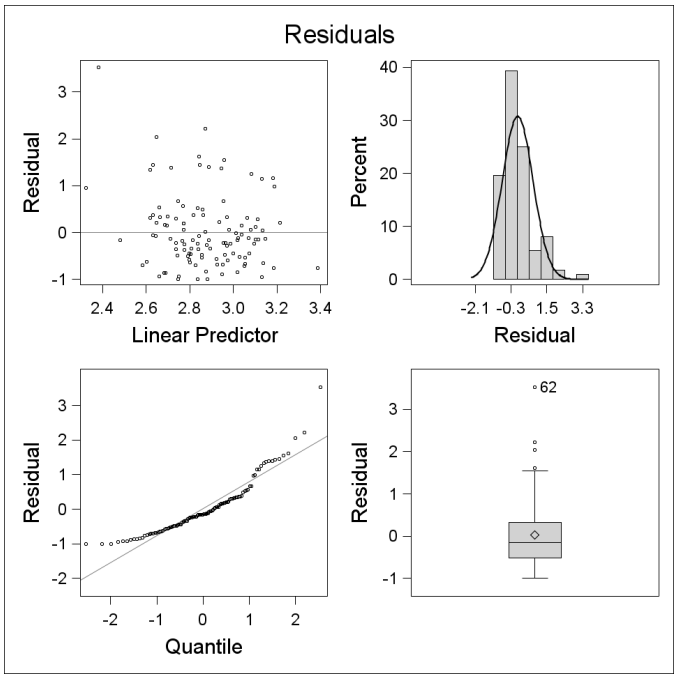


Figure 4.3: Pearson residuals panel for model (4.1).

$$\begin{aligned} txt.comp_i &\sim \text{Binomial}(n_i, \pi_i), \\ \text{logit}\,\pi_i &= \beta_0 + \beta_1(ctx.obs_{[1]i}) + \beta_2(dem.sex_{[1]i}) \\ &\quad + \beta_3(dem.sex_{[1]} \times ctx.obs_{[1]})_i \\ \text{Var}[txt.comp_i] &= \phi\pi_i(1-\pi_i)/n_i, \\ i &\in \{1, 2, \ldots, 112\}, \end{aligned} \tag{4.2}$$

where n is the number of questions answered, π is the probability of a correct answer, and i is the subject index. I assumed that the distribution of text comprehension is Binomial because it arises from a set of n_i independent Bernoulli trials (answers to forced multiple choice questions), each having the probability of success equal to π_i. There was a substantial underdispersion in the data ($\sigma^2_{pr} = 0.35$) which I addressed by the including the multiplicative scale parameter ϕ ($\sigma^2_{pr} = 1.00$). The residuals panel (Figure 4.4) shows a good fit of the model.

Results

The results of fitting the number of mindless reading episodes model (4.1) are shown in Table 4.4. The observer variable is not significant ($p = .7366$) but its interaction with subject's sex, while not reliable, ascends to the rank of a trend ($p = .0776$).

The results of fitting the text comprehension model (4.2) are shown in Table 4.4. Here, neither the observer effect ($p = .6280$) nor the interesting interaction ($p = .2257$) are significant.

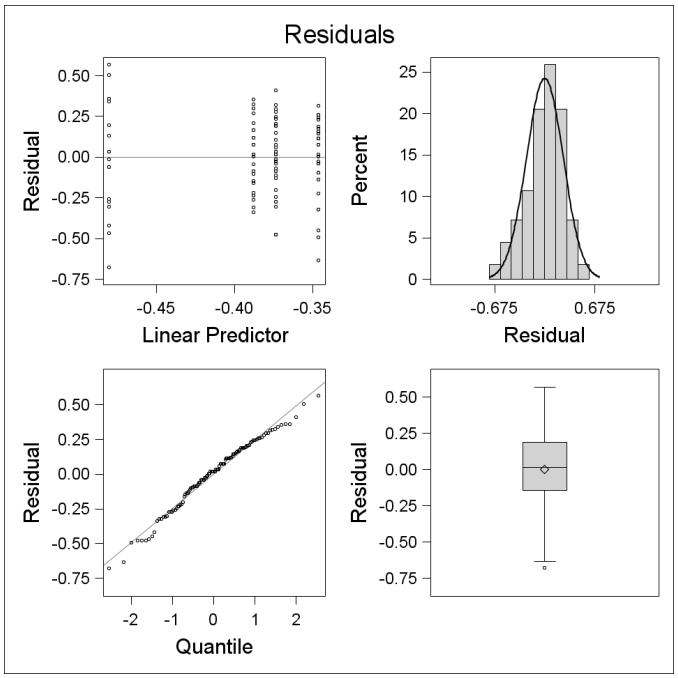


Figure 4.4: Pearson residuals panel for model (4.2).

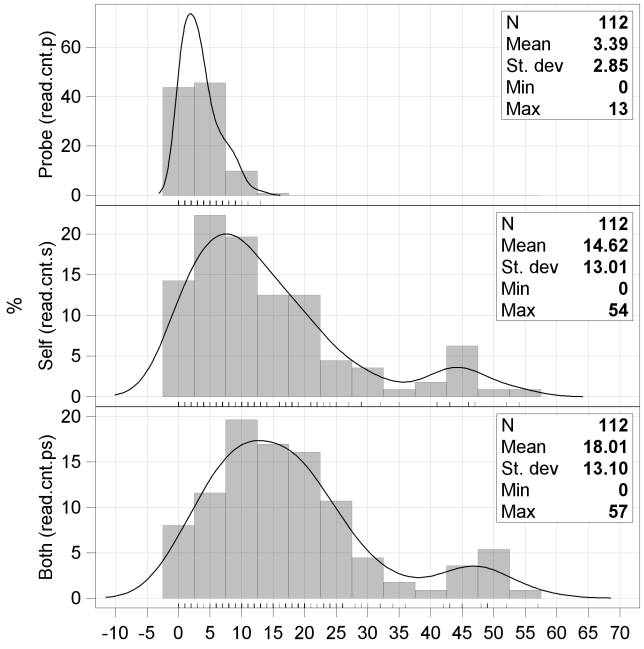


Figure 4.5: Histogram of the number of P (*read.cnt.p*), S (*read.cnt.s*), and P+S (*read.cnt.ps*) mindless reading episodes.

Table 4.4: Parameter estimates for model (4.1).

Effect	β	SE	DF	t	p
Intercept	−1.3437	0.1184	108	−11.35	<.0001
ctx.obs [1]	−0.06115	0.1814	108	−0.34	.7366
dem.sex [1]	−0.2656	0.1972	108	−1.35	.1808
ctx.obs×dem.sex	0.5130	0.2879	108	1.78	.0776
ϕ	9.8718				

Conclusions and Discussion

The current data does not contain evidence of the observer effect. Furthermore, I failed to establish a statistically reliable interaction between the gender of the experimenter and that of the subject which was reported earlier by Singer (1988). For these reasons, I exclude the observer variable (*ctx.obs*) from all subsequent analyses.

At the same time, I want to emphasize that my results are not entirely at odds with those of Singer (1988) in that while I was not able to corroborate his finding, I found a tendency in the mentioned interaction that should be studied further.

Table 4.5: Parameter estimates for model (4.2).

Effect	β	SE	DF	t	p
Intercept	−0.3739	0.03964	108	−9.43	<.0001
ctx.obs [1]	0.02743	0.05644	108	0.49	.6280
dem.sex [1]	−0.01404	0.06201	108	−0.23	.8213
ctx.obs×dem.sex	−1202	0.09867	108	−1.22	.2257
ϕ	0.1247				

4.4 Propensity to Lapse into Mindless Reading

Motivation

Figure 4.5 shows distributions of the number of P (*read.cnt.p*), S (*read.cnt.s*), and both (*read.cnt.ps*) mindless reading events. The distribution of *read.cnt.ps* is bimodal with most of the probability mass around the first local maximum. This dual modality may indicate that the distribution arises as a result of two interleaved stochastic processes. Furthermore, it is clear that S events contribute most of the probability mass, including the second local maximum.

To accommodate the bimodal nature of the *read.cnt.ps* distribution I performed two analysis. In Section 4.4.1, I look at cases of low self-reporters only (i.e., I exclude the 11 high self-reporters). It is imperative that the reader understands the rationale behind this course of action. The Poisson distribution cannot account for two modes and therefore the Poisson regressions model that I intend to use is inappropriate unless observations contributing one of the modes are dropped. In Section 4.4.2, I explore the possibility of two processes which result in low and high self-report numbers.

Note, that combining the two types of mindless reading events gives rise to a distribution resembling more the Poisson, which describes the count of independent rare events. That fact alone can be seen as an indication that both the P and S events need to be taken into account together to properly sample the incidence of mindless reading episodes.

4.4.1 Number of Mindless Reading Episodes

Models and Data

To investigate the effect that subject-level variables had on the propensity to lapse into mindless reading (P and S events together) given that only the low self-reporters are considered, I fitted the following model

$$
\begin{aligned}
& read.cnt.ps_i \leq 40, \\
& read.cnt.ps_i \sim \text{Poisson}(\lambda_i), \\
& \log\ \lambda_i = \beta_0 + \beta_1(txt.rom_{[0]i}) + \beta_2(txt.dra_{[0]i}) \\
& \qquad + \beta_3(txt.int_i) + \beta_4(read.spd_i) \\
& \qquad + \beta_5(id.wm.fl_i) + \beta_6(id.sat_i) + \beta_7(ctx.fat_i) \\
& \qquad + \beta_8(ctx.pre_i) + \beta_9(ctx.crv.3s_{[0]i}) \\
& \qquad + \beta_{10}(ctx.crv.3s_{[1]i}) + \beta_{11}(ctx.totd.3s_{[1]i}) \\
& \qquad + \beta_{12}(ctx.totd.3s_{[2]i}) + \log(read.t.tot_i/10), \\
& \text{Var}[read.cnt.ps_i] = \phi\lambda_i, \\
& \qquad i \in \{1, 2, \ldots, 92\},
\end{aligned}
\tag{4.3}
$$

where i is the subject index. Of the total 112 subjects, 101 had less then 40 mindless reading episodes during the experiment, but only 92 of them had no missing values in any of the variables included. I used time spend reading in minutes ($read.t.tot$) as an offset variable and divided it by 10 to report rates of mindless reading per 10-minute time interval. I also included a multiplicative scale parameter ϕ on the variance function to account for overdispersion. Good variance model was warranted by $\sigma^2_{PR} = 1.01$. The residuals panel (Figure 4.6) indicates that the model fits the data well.

Results

The results of fitting model (4.3) are shown in Table 4.6. Interestingness of text ($txt.int$) is the first significant covariate ($p = .0057$). A unit increase in that interestingness (measured on a scale of one to seven) is expected to decrease the rate of mindless reading per 10 minutes by about 0.88. That is, being interested in what is being read was conducive to reading mindfully ($\rho_{read.cnt.ps,txt.int} = -.31$). The estimated effect of interestingness of text is shown in Figure 4.7.

Fatigue ($ctx.fat$) is the second significant covariate ($p = .0174$). A unit increase in fatigue (measured on a scale of one to seven) yields an increase of one mindless reading episode per 10 minutes. Hence, being fatigued was incompatible with efficient reading ($\rho_{read.cnt.ps,ctx.fat} = .24$). The estimated effect of fatigue is shown in Figure 4.8.

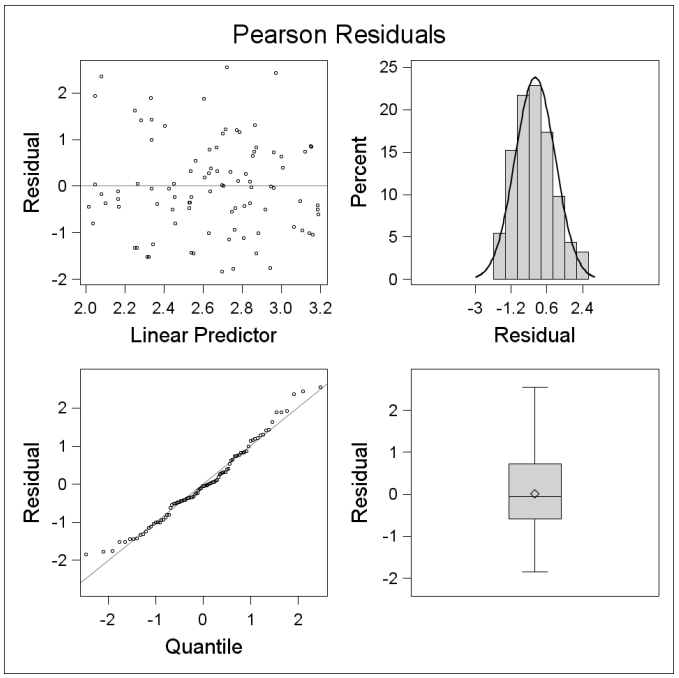


Figure 4.6: Pearson residuals panel for model (4.3).

4.4.2 Low and High Self-Reporters

As I have indicated earlier, the bimodal profile of the distribution of the number of mindless reading episodes ($read.cnt.ps$; Figure 4.5) may indicate that two stochastic processes generated these counts. Most of the mass of the distribution of $read.cnt.ps$ comes from self-reported mindless reading ($read.cnt.s$). That includes the second local maximum. Therefore, if two processes were indeed responsible for generating the observed data, then the second process must be indigenous to self-caught mindless reading. That is why in this subsection I focus only on self-reported mindless reading and attempt to discover factors associated with being a high self-reporter.

Only a small portion of subjects in the current experiment were high self-reporters. If the variability in the population of readers has been accurately represented in my sample, then roughly 1 in 10 people should be expected to score high on the scale of self-reported mindless reading. Of course, there is no way of knowing if these high self-reporters would still be prone to noticing their mindless reading more if they did not participate in an eye-tracking experiment which is hardly a "just another day" type of an experience. This issue should be investigated further in larger experiments.

Table 4.6: Parameter estimates for model (4.3). Levels of classification variables in square brackets.

Effect	β	SE	DF	t	p
Intercept	0.7959	0.5903	79	1.35	.1814
txt.rom [0]	−0.1101	0.1350	79	−0.82	.4173
txt.dra [0]	−0.1988	0.1805	79	−1.10	.2740
txt.int	−0.1248	0.04394	79	−2.84	**.0057***
read.spd	0.001549	0.001122	79	1.38	.1712
id.wm.fl	0.08730	0.3692	79	0.24	.8137
id.sat	−0.2205	0.5893	79	−0.37	.7093
ctx.fat	0.1003	0.04129	79	2.43	**.0174***
ctx.pre	−0.02339	0.02332	79	−1.00	.3191
ctx.crv.3s [0]	−0.2991	0.1521	79	−1.97	.0528
ctx.crv.3s [1]	−0.09273	0.1515	79	−0.61	.5422
ctx.time.3s [1]	0.3418	0.2120	79	1.61	.1109
ctx.time.3s [2]	0.2797	0.2157	79	1.30	.1986
ϕ	4.3793				

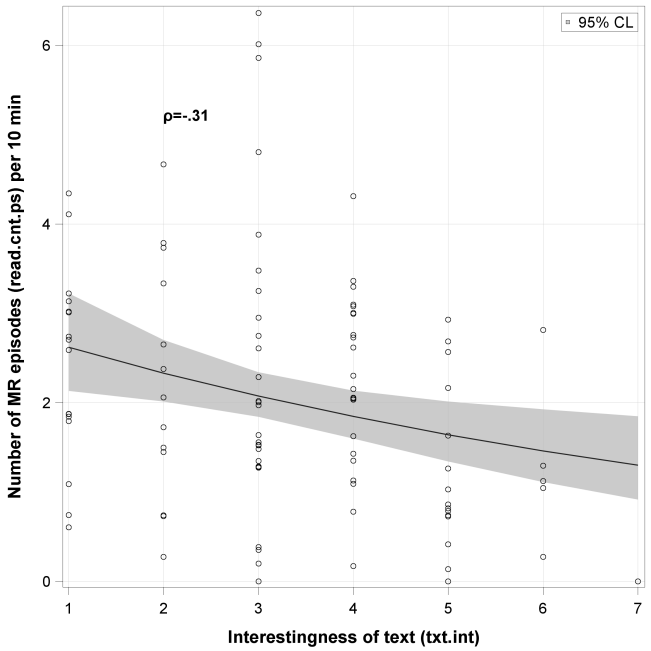


Figure 4.7: Estimated effect and observed values of interestingness of text ($txt.int$) against the number of mindless reading episodes ($read.cnt.ps$).

Models and Data

The distribution of $read.cnt.ps$ is bimodal and most likely a mixture of two Poisson distributions. Based on the current data, about 70 minutes of reading result in the first of these distributions having the mean of about 15 and the second distribution having the mean of about 50 (Figure 4.5). Fitting a mixture of two Poisson distributions model would be ideal in this situation, but because I am interested in estimating 12 coefficients, it is not possible to fit such a large model to the current data. That is because while the parameters of the first distribution could be estimated reliably (101 observations), the same cannot be said about the second distribution (only 11 observations).

To circumvent this problem, I divided the subjects into two groups: Low- and high self-reporters. I fitted the following model to check which factors could influence the probability of being in the second group

$$
\begin{aligned}
read.cnt.s.bin_i &= \begin{cases} 0 & \text{if } read.cnt.s_i < 40 \\ 1 & \text{otherwise} \end{cases}, \\
read.cnt.s.bin_i &\sim \text{Bernoulli}(\pi_i), \\
\text{logit}\,\pi_i &= \beta_0 + \beta_1(txt.rom_{[0]i}) + \beta_2(txt.dra_{[0]i}) \\
&\quad + \beta_3(txt.int_i) + \beta_4(read.spd_i) \\
&\quad + \beta_5(id.wm.fl_i) + \beta_6(id.sat_i) + \beta_7(ctx.fat_i) \\
&\quad + \beta_8(ctx.pre_i) + \beta_9(ctx.crv.3s_{[0]i}) \\
&\quad + \beta_{10}(ctx.crv.3s_{[1]i}) + \beta_{11}(ctx.totd.3s_{[1]i}) \\
&\quad + \beta_{12}(ctx.totd.3s_{[2]i}) + \log(read.t.tot_i/10), \\
\text{Var}[read.cnt.s.bin_i] &= \phi\pi_i(1-\pi_i), \\
i &\in \{1, 2, \ldots, 99\},
\end{aligned}
\tag{4.4}
$$

where π is the probability of being a high self-reporter, and i is the subject index. Of the total 112 subjects, 99 had no missing values in any of the variables included. The multiplicative scale parameter on the variance function allowed for better variance modeling ($\sigma^2_{pr} = 1.00$) as compared to the model without that parameter ($\sigma^2_{pr} = 0.74$).

Results

The results of fitting model (4.4) are shown in Table 4.7. Reading speed is the first statistically significant covariate ($p = .0133$) and it correlates positively with being a high self-reporter of mindless reading ($\rho_{read.cnt.s.bin,read.spd} = .24$). Quantitatively, a unit increase in reading speed (measured in words per minute) translates into an increase of about 1% in the odds of being a high self-reporter. It appear then that readers who devoured the text at higher rates were also more likely to catch themselves reading mindlessly. The estimated effect of reading speed is shown in Figure 4.9.

Preoccupation ($ctx.pre$) is the second significant covariate ($p = .0226$). Being preoccupied was associated with a higher probability of being high self-reporter ($\rho_{read.cnt.s.bin,ctx.pre} = .24$). More specifically, the model indicates that a unit increase in preoccupation (measured on the scale of 0 to 21) is expected to increase the odds of reporting over 40 mindless reading episodes by about 37%. The estimated effect of preoccupation is shown in Figure 4.10.

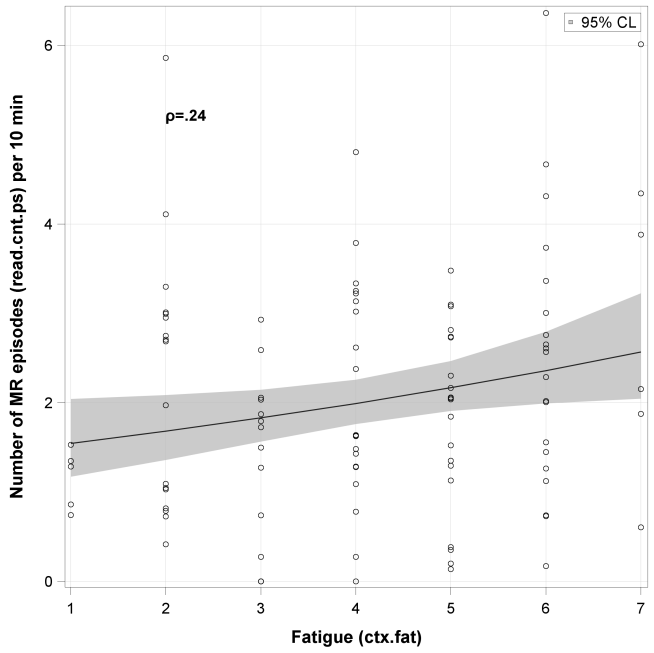


Figure 4.8: Estimated effect and observed values of fatigue ($ctx.fat$) against the number of mindless reading episodes ($read.cnt.ps$).

4.4.3 Conclusions and Discussion

The distribution of the number of mindless reading episodes (i.e., both probe- and self-caught together) shows a bimodal profile. My investigation of the first mode of that distribution revealed that higher levels of fatigue increased the rate at which mindless reading was experienced by the subjects but that interestingness of text had the opposite effect. These results are consistent with prior research. For example, Kane et al. (2007) and Teasdale et al. (1995) found fatigue responsible for a significant increase in the rate of mindless reading episodes. Additionally, Grodsky & Giambra (1989) employed a five-point Likert-scale measurement[1] similar to the one I used to retrospectively sample interest of passages of text read by their subjects. They found that the frequency of mind-wandering decreased with greater interestingness of text. More recently, Smallwood, Nind, & O'Connor (2009) used that same measure and found a similar relationship.

I approached the investigation of the bimodal nature of the number of mindless reading episodes distribution by dividing the subjects into two groups. Because the second mode was contributed entirely by self-caught mindless reading, which group a subject belonged to depended only on the number of self-reported mindless reading episodes. Subjects with no-more-than 40 episodes of self-caught mindless reading were classified as low self-reporters; the rest were high self-reporters. I have found that higher levels of preoccupation made the subjects more likely to be in the high self-reporter group, a results noticed earlier by Schacter (2001).

Furthermore, I have found that faster-reading subjects were more likely to be in the high self-reporter group. This result has no precedence in the literature and I investigate it in more detail in the next section.

[1] "On a scale of 0 to 4 how interesting did you find the text?" (0 – "I was NOT interested in this material at all and did not enjoy reading it"; 4 – "This is the most interesting material I've read in the past year and I would like to read even more on this topic").

Table 4.7: Parameter estimates for model (4.4). Levels of classification variables in square brackets.

Effect	β	SE	DF	t	p
Intercept	−8.1515	4.1024	86	−1.99	.0501
txt.rom [0]	1.3684	1.1046	86	1.24	.2188
txt.dra [0]	−0.5761	1.1663	86	−0.49	.6226
txt.int	−0.2292	0.2921	86	−0.78	.4349
read.spd	0.01451	0.005737	86	2.53	**.0133***
id.wm.fl	−3.4284	2.3914	86	−1.43	.1553
id.sat	−1.9052	3.6228	86	−0.53	.6003
ctx.fat	0.09120	0.2596	86	0.35	.7262
ctx.pre	0.3117	0.1343	86	2.32	**.0226***
ctx.crv.3s [1]	−0.8080	1.0542	86	−0.77	.4455
ctx.crv.3s [0]	−0.4388	0.9073	86	−0.48	.6299
ctx.time.3s [2]	1.1964	1.3032	86	0.92	.3612
ctx.time.3s [1]	2.1004	1.1384	86	1.85	.0685
ϕ	0.5577				

4.5 Reading speed

Motivation

In the previous subsection, I have indicated that in the current experiment faster readers experienced more mindless reading episodes. One plausible explanations of this result is that if reading speed is a proxy for reading skill, then reading in general should be easier for faster readers. This is important, because the incidence of mind-wandering has been shown to decrease in the context of cognitively demanding tasks (Smallwood & Schooler, 2006) and if that is also the case with reading then faster readers should be expected to mind-wander more.

However, because this result is correlational in nature, an alternative explanation exists. Namely, it could be that subjects who experienced more mindless reading episodes read at higher average speeds because of that. This would be consistent with the hypothesis that text processing is shallower during mindless reading as suggested earlier by Reichle, Reineberg, & Schooler (2010) and Schad, Nuthmann, & Engbert (2012) and as I show in this dissertation (Sections 4.6 and 4.10). More specifically, that hypothesis is based on the observation that the effects of lexical and linguistic word characteristics are at-

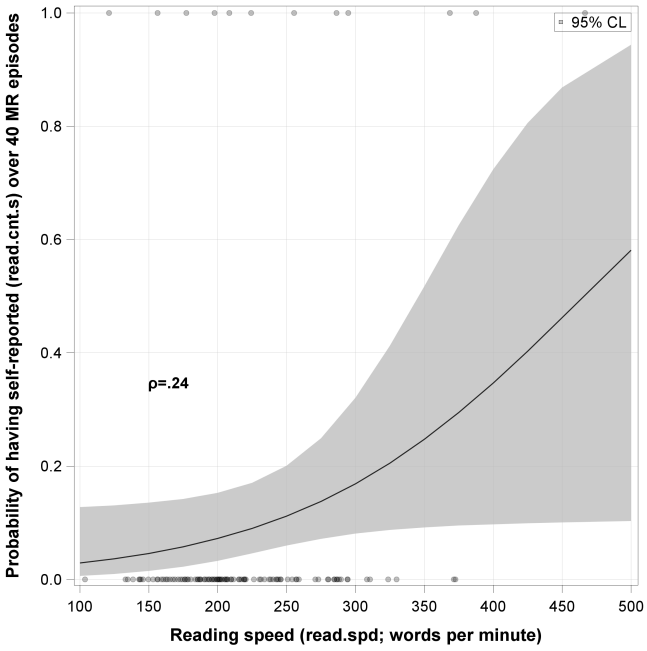


Figure 4.9: Estimated effect and observed values of reading speed ($read.spd$) against the probability of being in the group of high self-reporters of mindless reading ($read.cnt.s.bin$).

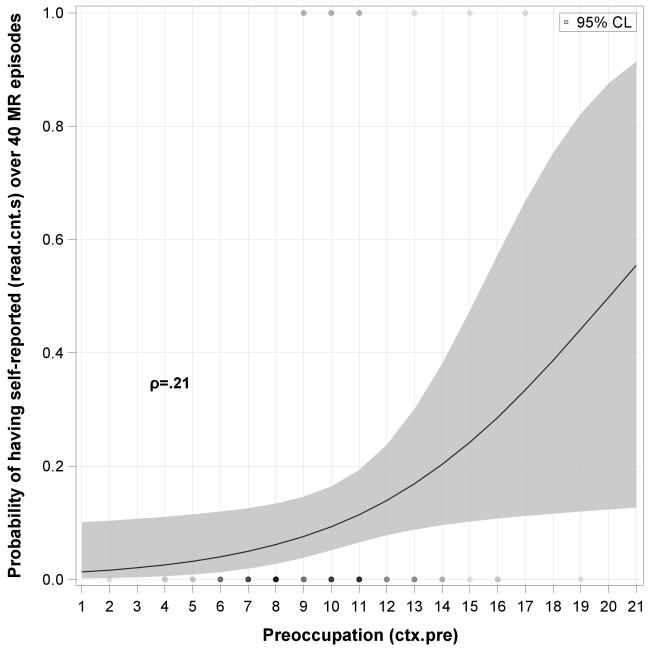


Figure 4.10: Estimated effect and observed values of preoccupation ($ctx.pre$) against the probability of being in the group of high self-reporters of mindless reading ($read.cnt.s.bin$).

tenuated when a mind of a reader has wandered. That is, as eye movements are programmed, the *where* decision (which depends largely on low-level visual factors) remains in effect while the *when* decision (which depends largely on lexical and linguistic properties of words) intervenes to a lesser degree or not at all (for a review of the when-and-where, see Rayner, 1998). Importantly, Rayner & McConkie (1976) has shown that these two decisions can in fact be made somewhat independently.

Finally, it is possible that the combination of the two explanations I have offered above determines the interaction between reading speed and propensity to mind-wander. In this section, however, I investigate the second of the two explanations only. I do that by checking if reading speeds during normal and mindless reading differ. If mindless reading proceeds faster than normal reading then that would support the hypothesis that text processing during mindless reading is impoverished.

Models and Data

To check if the mean reading speed was different in normal and mindless reading, I fitted a series of 13 identical models (one per time window), each given as

$$\begin{aligned} read.spd.r_{ij}|s_i &\sim \text{Normal}(\mu_{ij}, \sigma^2_{ij}), \\ \mu_{ij} &= \beta_0 + \beta_1(r_{ij}) + s_i, \\ \text{Var}[read.spd.r_{ij}] &= \sigma^2_{ij}, \\ s_i &\sim iid\ \text{Normal}(0, \sigma^2_s), \\ i &\in \{1, 2, \ldots, 77\}, \\ j &\in \{1, 2, 3\}, \end{aligned} \tag{4.5}$$

where μ is the mean being modeled, s is the subject random effect, i is the subject index, j is the reading mode (i.e., N, P, and S) index.

Results

Figure 4.11 shows the results of fitting model (4.5). The first row of the figure shows the mean reading speeds in normal reading (N), probe-caught mindless reading (P), and self-caught mindless reading (S). The second row shows the differences between these three means and the mean for normal reading; that makes it easy to see how mindless reading diverges from attentive reading. The three subsequent rows show p-values associated with the three comparisons between then N, P, and S conditions (i.e., N-P, N-S, and P-S). Finally, the two last rows report model diagnostic information in the form of the variance of Pearson residuals (ρ^2_{pr}; values close to one indicate no dispersion problems) and the root mean square error

(RMSE).

Because I use this figure template in several other analyses I discuss in the remainder of this dissertation I encourage the reader to understand it before moving on. I digress further to elucidate how I interpret this and other figures which report *p*-values for a number of time windows. Instead of cherry-picking statistically significant differences I require significant differences to be stable across several adjacent time windows. Additionally, I exercise even greater caution in interpreting results from the narrowest time windows because they are based on the smallest amount of data.

As is evident from the figure (rows 2–5), when reading mindfully, readers proceeded at 7.5–12.5 words per minute slower compared to the two mindless reading conditions. However, while this was the case for both probe-caught and self-caught mindless reading, only the self-caught mindless reading difference was statistically reliable and only in time windows of 15–40s. Because the number of probe-caught mindless reading episodes was substantially smaller (roughly half the amount) than self-caught mindless reading, this lack of statistical significance involving probe-caught mindless reading could be a result of insufficient statistical power, but there is no way of being sure.

The first row of the figure looks rather suspicious because it appears that reading speed is dropping proportionally to the time window size. However, this is not what the figure actually shows. Instead, this drop is associated with the intrinsic difficulty in calculating reading speed in the context of natural reading of ecologically valid text. The formula I used to calculate that speed was

$$read.spd.r = \frac{number.of.words.seen}{time.window.size} \cdot 60.$$

Clearly, the more words a reader sees and the narrower the time window the higher the reading speed. However, re-reading becomes more likely with the increase in time window size. That is, while the denominator of this formula increases proportionally to the time window size, the nominator does not increase nearly as fast because re-reading involves words which have already been seen (i.e., re-reading a word does not result in a larger nominator). That is why reading speed appears to drop in the figure. However, because of the relative stability of (1) the differences between normal reading and both kinds of mindless reading and (2) the differences between the two kinds of mindless reading themselves, it seems reasonable to assume that the formula I used captures a stable reading speed as modulated by re-reading.

Conclusions and Discussion

I have found reading speed to be smaller during normal reading as compared to self-caught mindless reading in time windows of 15–40s. The difference between the two speeds was about 10 words per minute. The fact that readers sped up while being mindless could plausibly be a result of the more superficial text processing, one in which the effects of perceptual word variables (e.g., word length) are used to program eye movements while the effects of lexical word variables (e.g., word frequency) are attenuated.

4.6 Text Comprehension

Motivation

To foreshadow results I present below, the current data indicates that reading mindlessly is incompatible with achieving good text comprehension. Although several prior studies (Reichle, Reineberg, & Schooler, 2010; Sayette, Reichle, & Schooler, 2009; Sayette, Schooler, & Reichle, 2009; Schad, Nuthmann, & Engbert, 2012; Schooler, Reichle, & Halpern , 2004) have documented similar associations, it remains unclear whether comprehension difficulty causes mind-wandering or whether mind-wandering causes comprehension difficulty. By one account, mind-wandering might cause readers to process text at a superficial level and thereby prevent them from constructing an accurate discourse model. However, by the other account comprehension difficulties may make readers disengage from text and lead to mind-wandering. In this section, I investigate these two possibilities and argue that because the text the subjects read was easy it is unlikely that difficulty understanding what was read induced mind-wandering and that the reverse is in fact what must have happened.

That the reading task was easy for the subjects plays a pivotal role in my argument because if I could show that bouts of mind-wandering are associated with impaired reading performance but happen in circumstances otherwise conducive to comprehension, then that would provide evidence for mindless reading being the cause of impaired reading comprehension and not the other way around. The novel the subjects read has been rated 11.9 on the Flesh-Kincaid grade-level index (Kincaid et al., 1975) so it was well within the reading ability of the subjects ($\mu_{dem.age} = 18.73$ years). Moreover, working memory tends to discriminate subjects with respect to performance, i.e., low-spans have problems with more difficult tasks compared to high-spans. Because as I show later no such discrimination was apparent in the current data, I see that as further evidence that the task was easy for everyone. This argument can be expressed using propositional logic: If P denotes "task is demanding" and Q denotes "working-memory capacity is associated with mind-wandering," then by *modus tollens* (i.e., the following argument: if P then Q; $\sim$Q; therefore $\sim$P) we can assert that the reading task was not demanding ($\sim$P) if we

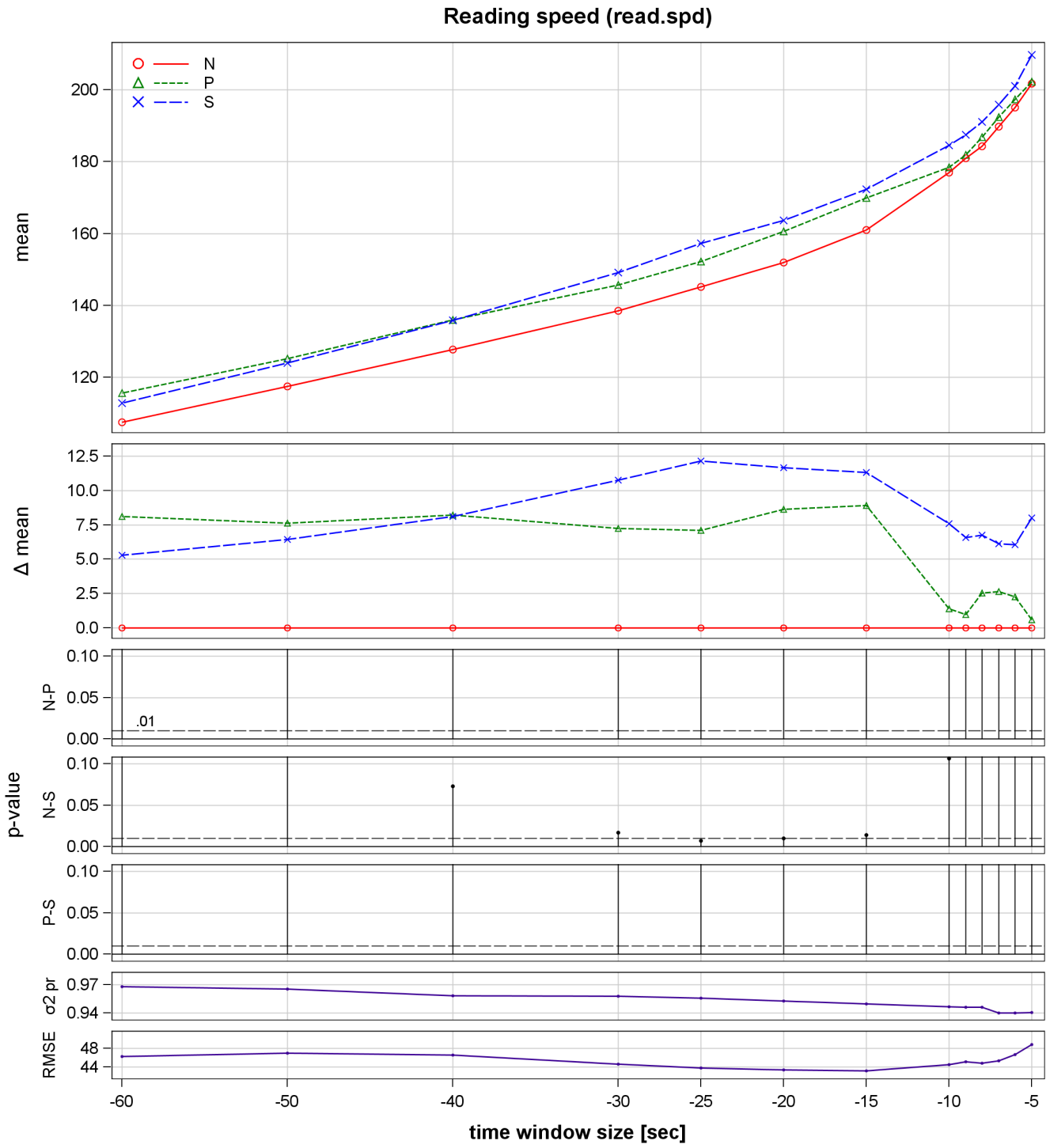


Figure 4.11: Reading speed (*read.spd*; in words per minute) for normal reading (N), probe-caught mindless reading (P), and self-caught mindless reading (S). Based on fitting model (4.5).

do not observe an association between working memory capacity and mind-wandering (∼Q).

Finally, the influence of working memory capacity on reading has been studied and it is known to directly affect comprehension by delimiting the amount of linguistic information that can be concurrently maintained and processed (Just et al., 1996; Miyake, Just, & Carpenter, 1994). Below, I show that this relationship is also present in my

data.

To summarize, in this section I test the following hypotheses:

- H1: Working-memory capacity is positively related to text comprehension during intervals of normal (i.e., attentive) reading (i.e., high-capacity readers should answer more of comprehension questions correctly than their low-capacity counterparts).
- H2: The rate of mindless-reading episodes is inversely related to comprehension (i.e., readers who frequently lapse into mind-wandering should answer fewer comprehension questions correctly than readers who less frequently succumb to mind-wandering).
- H3: Working-memory capacity is not related to the incidence of mindless-reading episodes (i.e., there should be no difference between how frequently low- and high-capacity subjects lapse into mind- wandering).

Models and Data

To test hypotheses H1 and H2, i.e., to investigate the influence that working memory capacity ($id.wm.fl$), and the number of normal reading ($read.cnt.n$) and mindless reading ($read.cnt.p$, $read.cnt.s$, and $read.cnt.ps$) episodes had on text comprehension ($txt.comp$), I fitted the following model

$$\begin{aligned} txt.comp_i &\sim \text{Binomial}(n_i, \pi_i), \\ \text{logit}\,\pi_i &= \beta_0 + \beta_1(id.wm.fl_i) + \beta_2(read.cnt.ps_i) \\ &\quad + \beta_3(id.wm.fl_i \times read.cnt.ps_i), \\ \text{Var}[txt.comp_i] &= \phi\pi_i(1-\pi_i)/n_i, \\ i &\in \{1, 2, \ldots, 112\}, \end{aligned} \tag{4.6}$$

where n is the number of questions answered, π is the probability of a correct answer, and i is the subject index. I assumed that the distribution of text comprehension is Binomial because it arises from a set of n_i independent Bernoulli trials (answers to forced multiple choice questions), each having the probability of success equal to π_i.

To test hypothesis H3, i.e., to check if working memory capacity ($id.wm.fl$) influenced the number of mindless reading episodes ($read.cnt.ps$), I fitted the following model

$$\begin{aligned} read.cnt.ps_i &\sim \text{Poisson}(\lambda_i), \\ \log\lambda_i &= \beta_0 + \beta_1(id.wm.fl_i) + \log(read.t.m_i/10), \\ \text{Var}[read.cnt.ps_i] &= \phi\lambda_i, \\ i &\in \{1, 2, \ldots, 112\}, \end{aligned} \tag{4.7}$$

where i is the subject index. Note, that I used time spend reading ($read.t.m$) as an offset variable and included a multiplicative scale parameter ϕ on the variance function to account for overdispersion. I divided time spent reading by 10 to report rates of mindless reading per 10-minute time interval.

Results

I fitted model (4.6) four times to accommodate the four types of reading episodes: (a) normal reading, (b) probe-caught mindless reading, (c) self-caught mindless reading, and (d) total mindless reading (i.e., probe- and self-caught episodes combined). The results are shown in Table 4.8 and on Figure 4.12. Consistent with hypothesis H1, I found working memory capacity to be positively correlated with text comprehension: An increase from the (hypothetical) minimum (i.e., zero) to maximum (i.e., one) in working-memory capacity was expected to increase the odds of answering a comprehension question correctly by 163% (Plots 1; $p \leq .0073$). Consistent with hypothesis H2, the number of normal reading episodes was positively correlated with comprehension (Plot 2a; $p < .0001$), but the number of self- caught and total mindless-reading episodes were negatively correlated with comprehension (Plots 2c and 2d; $p \leq .0004$). However, the number of probe-caught mindless-reading episodes was not associated with comprehension (Plot 2b; $p = .2904$). This could be due to the lack of statistical power as the average number of probe-caught episodes was only 3.4 ($\sigma = 2.85$); for comparison, I recorded an average of 14.62 ($\sigma = 13.01$) self-caught episodes. Working-memory capacity did not mediate the effect of neither the number of normal nor the number of mindless reading episodes ($p \geq .7150$).

To test hypothesis H3, I examined whether low- and high-capacity participants differed in their rates of mindless reading. I did that by fitting model (4.7) three times to relate working-memory capacity to the rate of (a) probe-caught mindless reading, (b) self-caught mindless reading, and (c) total mindless reading. I found no significant relationships ($p \geq .7820$). This indicates that, consistently with hypothesis H3, working-memory capacity was not predictive of the frequency with which the subjects lapsed into mind-wandering.

Conclusions and Discussion

Although working-memory capacity was positively related to comprehension and mind-wandering was inversely related to comprehension, working memory capacity was not related to the propensity to mind-wandering. Altogether, these results suggest that mind-wandering disrupts comprehension of otherwise easy-to-understand text.

The Balance between Mindfulness and Mind-Wandering, pg 9, p 1. They show that the difficulty of an OSPAN task (i.e., list length) did not influence the likelihood of mind-wandering which suggests that the

Table 4.8: Parameter estimates for model (4.6).

ET[1)]	Effect	β	SE	DF	t	p
	Intercept	0.7828	0.0640	108	12.23	<.0001
	id.wm.fl	0.9682	0.3308	108	2.93	**.0042**
N	**read.cnt.ps**	0.0553	0.0101	108	5.49	**<.0001**
	Interaction	0.0205	0.0560	108	0.37	.7150
	ϕ	0.0947				
	Intercept	0.7622	0.0710	108	10.74	<.0001
	id.wm.fl	1.0062	0.3682	108	2.73	**.0.0073**
P	read.cnt.ps	−0.0264	0.0248	108	−1.06	.2904
	Interaction	0.0403	0.1406	108	0.29	.7750
	ϕ	0.1211				
	Intercept	0.7709	0.0679	108	11.35	<.0001
	id.wm.fl	1.0232	0.3514	108	2.91	**.0044**
S	**read.cnt.ps**	−0.0187	0.0051	108	−3.68	**.0004**
	Interaction	−0.0092	0.0275	108	−0.33	.7387
	ϕ	0.1090				
	Intercept	0.7723	0.0674	108	11.45	<.0001
	id.wm.fl	1.0221	0.3486	108	2.93	**.0041**
P+S	**read.cnt.ps**	−0.0198	0.0050	108	−3.93	**.0001**
	Interaction	−0.0097	0.0280	108	−0.35	.7294
	ϕ	0.1071				

[1)] Episode type: N–normal reading, P–probe caught MR, S–self-caught MR

negative performance-WMC correlation was not due to low-WMC subjects failing to remain engaged because of having difficulty with the task. Instead, they found that performing poorly on a trial was associated with less mind-wandering on the subsequent trial.

4.7 Eye-Movement Variables

Beginning with this section, I move away from subject-level variables and move towards eye-movement (and word-level) variables which play pivotal role in my attempts to detect mindless reading. Table 4.9 lists all eye-movement variables I investigate in the remainder of this dissertation.

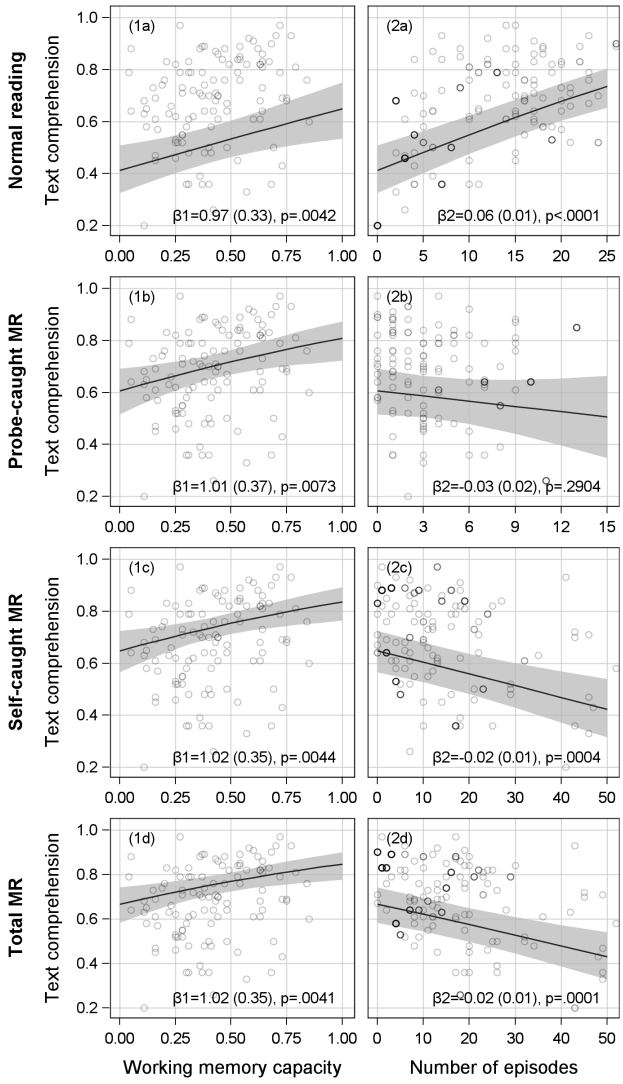


Figure 4.12: The influence of working memory capacity ($id.wm.fl$) and the number of normal ($read.cnt.n$) and mindless ($read.cnt.p$, $read.cnt.s$, and $read.cnt.ps$) reading episodes on text comprehension ($txt.comp$; model (4.6)).

4.8 Off-Screen Fixations

Motivation

Mindless reading could result in processing of text shallow enough to make readers unaware of the boundaries of the screen. That is, a mindless reader could continue "reading" past the right edge of the screen or even make off-screen fixations going in other directions (which could result from overshot return saccades). In fact, Reichle, Reineberg, & Schooler (2010) reported a higher incidence of off-screen fixations associated with self-caught mindless reading. That higher incidence was especially visible prior to subjects catching themselves on being mindless. Following up on this earlier report, in this section, I check if mindless reading which happened during the current experiment resulted in a number of off-screen fixations larger than that expected to happen during normal reading.

In order to better understand what happened during my experiment, I plotted the observed probabilities of subjects making an off-screen fixation for seven time window sizes and broke it down into the four cardinal directions (East is to the right of the screen, i.e., 90°; Figure 4.13). The probabilities of a fixation happening in any of the corners (e.g., North-West) were very small and

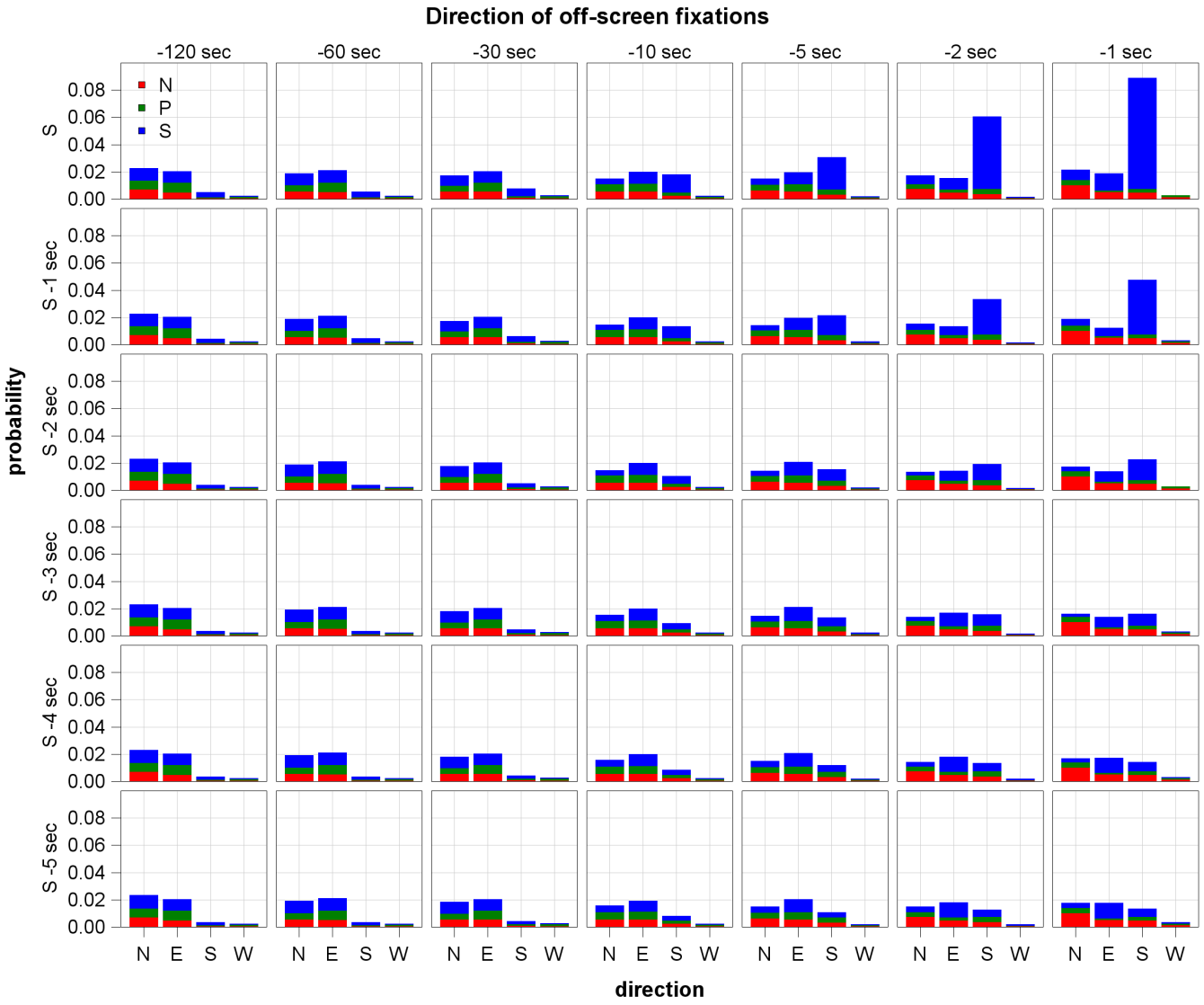


Figure 4.13: Probability of an off-screen fixation ($em.posf$) going in one of the four cardinal directions (North, East, South, and West; East is to the right of the screen, i.e., 90°) for normal reading (N), mindless probe-caught reading (P), and mindless self-caught reading (S). The columns show time windows that include an increasing number of off-screen fixations (from the right: 1 to 120 seconds prior to the onset of a thought-sampling probe or self-report of mindless reading). The rows show how many seconds of off-screen fixations going South were excluded from the bar plot data for the S condition (from the top: 0 to 5 seconds). Probabilities for the N and P conditions are the same for all rows; only those for the S condition change.

therefore I do not show them. From the first row, it is evident that a vast majority of fixations made immediately prior to self-reporting mindless reading landed below the screen (i.e., in the South direction). Because subjects needed to find and press the Z key on a keyboard located on the table in front of them to self-report mindless reading, this high incidence of below-the-screen fixations must have been, at least in part, an artifact of the experimental setup. Consequently, it seems prudent to exclude all off-screen fixations going in that direction before comparing normal and mindless reading to prevent the need to look down from contaminating the results.

Models and Data

To check if the probability of making an off-screen fixation ($em.posf$) was different in normal and mindless reading, I fitted a series of 19 identical models (one per time win-

Table 4.9: Eye-movement variables I investigate in this dissertation.

Long name	Short name	FP[1)]
Inspection durations [ms]		
First-pass fixation duration	$em.fpfd$	+
Second-pass fixation duration	$em.spfd$	
Gaze duration	$em.gd$	+
Single fixation duration	$em.sfd$	+
First-fixation duration	$em.ffd$	+
Inspection numbers		
Number of first-pass fixations	$em.nfpf$	+
Other numbers		
Number of blinks	$em.nb$	
Number of first-pass regressions	$em.nfpr$	+
Inspection probabilities		
Prob. of skipping	$em.p0$	+
Prob. of one fixation	$em.p1$	+
Prob. of two or more fixations	$em.p2$	+
Other probabilities		
Prob. of off-screen fixation	$em.posf$	
Prob. of regression	$em.pr$	+
Prob. of landing site left off-center	$em.plsloc$	+
Saccade amplitudes [char]		
Forward incoming saccade amp.	$em.fisa$	+
Forward outgoing saccade amp.	$em.fosa$	+
Regressive outgoing saccade amp.	$em.rosa$	+
Saccade other		
Landing site eccentricity	$em.lse$	+
Pupilometry [%]		
Pupil diameter	$em.pd$	

[1)] A plus indicates first-pass (FP) reading variables

dow), each given as

$$\begin{aligned} em.posf_{ij}|s_i &\sim \text{Binomial}(n_{ij}, \pi_{ij}), \\ \text{logit}\,\pi_{ij} &= \beta_0 + \beta_1(r_{ij}) + s_i, \\ \text{Var}[em.posf_{ij}] &= \phi\pi_{ij}(1-\pi_{ij})/n_{ij}, \\ s_i &\sim iid\ \text{Normal}(0, \sigma_s^2), \\ i &\in \{1, 2, \ldots, 99\}, \\ j &\in \{1, 2, 3\}, \end{aligned} \tag{4.8}$$

where π is the probability being modeled, n is the number of fixations, s is the subject random effect, i is the subject index, and j is the reading mode (i.e., N, P, and S) index. As I have indicated above, all below-the-screen fixations were excluded from this analysis.

Results

Figure 4.14 shows the results of fitting model (4.8). As is clear from the figure, there is no evidence for the difference between normal and self-caught mindless reading (the one significant p-value for time window of one second is most likely spurious). There is, however, some evidence that probe-caught mindless reading could result in a higher incidence of off-screen fixations than normal reading, as indicated by time windows of size 40–60 seconds. In these three time windows, the mean probability lines for the P and N conditions start diverging and the difference between them can be classified as a trend ($p < .1$).

Conclusions and Discussion

I have found weak evidence that 40–60 seconds prior to discovering a subject was unaware of having been reading mindlessly off-screen fixations were more probable than during normal reading. However, I have not detected any differences neither between normal reading and self-caught mindless reading nor between the two kinds of mindless reading. These results suggest that when focused on internal trains of thought a reader's attention may be so decoupled from the text being read that their visual focus fails to be constrained to the screen. These results further suggest that a reader can be engaged in mindless reading without being able to realize it for as long as one minute or more. Because adult readers read at an average speed of about 250 words per minute one minute is a fairly long time. Therefore, collectively, these two conclusions may indicate that long episodes of mindless reading are responsible for the biggest deficits in overall text comprehension.

Because of the experimental setup, the current data cannot be used to assess if mindless reading can produce a larger proportion of below-the-screen fixations. I cannot, however, think of anything that would make the lower edge of the screen different than the other edges. In fact, because reading English text commences from left to write, the bottom edge could be the least important. Employing a purely mouse-driven interface which would allow subjects to remain focused on the screen the entire time, even when responding to probes or self-reporting mindless reading, would be one way of address this methodological deficiency of the current experimental method.

Reichle et al. (2010) reported an increased incidence of off-screen fixations during self-caught mindless reading when compared to normal reading and probe-caught mindless reading. Because their experimental setup was almost identical to mine, the above conclusions naturally extend to include their study as well. Because I excluded below-the-screen fixations, this could also explain why I failed to corroborate their findings.

All that said, however, it is important to remember that the probability of making an off-screen fixation is very

small to start with irrespective of the attentive state of a reader; in the current data, depending on the time window size, it was typically smaller than .008. This has implications for detecting mindless reading. Namely, it does not appear that such rare event could on its own be harvested for evidence of mindless reading and consequently I do not use this variable in my modeling endeavors. In the future, it may be wise, however, to treat this event as a supplementary indicator of a reader's mindlessness.

Finally, as I have indicated, below-the-screen fixations during self-caught mindless reading are likely mostly due to the experimental setup. Indeed, as shown in Figure 4.13, the probability of these fixations occurring increases drastically immedately prior to the subject self-reporting mindless reading. By excluding between 1–5 seconds worth of these off-screen fixations I was able to establish what seems to be a "normal" probability of below-the-screen fixation, i.e., the probability that would be expected to happen if the experimental setup did not interfere. More specifically, excluding three seconds prior to a self-report of mindless reading appears to be enough to level things off. Therefore, the current data indicates that if our goal was to investigate the moment when a mindless reader is regaining their meta-conscious capacity, then it is that three-second time window that we should be looking into. However, because the primary goal of this dissertation is to detect mindless reading, in my modeling work I focus on eye-movements indigenous to that mindless reading only, i.e., those that happen before three seconds prior to a self-report (note that this implies exclusions to neither normal reading nor probe-caught mindless reading).

Note that I do not include any off-screen fixations in any of the subsequent analyses. In fact, in the remainder of this document I focus entirely on fixations made on text, i.e., I do not look at fixations made on a page unless they were made on an actual word (or, to be precise, on its bounding box; see Figure 3.1).

4.9 Extreme Fixation Durations

Motivation

In reading experiments, fixations shorter than about 80 ms or longer than about 1000 ms are typically treated as outliers (Inhoff & Radach, 1998; Liversedge, Paterson, Pickering, 1998). However, because little is known about eye-movements of mindless readers, some researchers (e.g., Reichle, Reineberg, & Schooler, 2010) decided to retain all fixations not to risk discarding potentially useful information.

In this section, I investigate the prevalence of extremely short and extremely long fixations (i.e., those falling outside of the 80–1000 ms duration range) in normal and mindless reading. Because this issue has not been investigated before and therefore there is no way of knowing whether or not excluding extreme fixations should be adopted as a practice in mindless reading studies. If both normal and mindless reading produce the same proportion of these extreme fixations, then that would lend support for discarding them. Conversely though, if text processing during mindless reading is disrupted to a point when extreme fixations creep in in numbers greater than during normal reading, then that would make them diagnostic of the decoupled attention characteristic of mind-wandering. Consequently, apart from indicating that extreme fixations should be retained, this would also mean that they could be helpful in detecting mindless reading. In the reminder of this section, I refer to the shorter-than-80, 80-1000, and longer-than-1000 ms fixations as *extremely short*, *normal*, and *extremely long*, respectively.

Models and Data

To check if there was a difference in the probability of extreme fixation occurring during normal and mindless reading, I fitted a series of 19 identical models (one per time window) first to extremely short fixation data and then to extremely long fixation data. Each model was given as

$$\begin{aligned} p_{ij}|s_i &\sim \text{Binomial}(n_{ij}, \pi_{ij}), \\ \text{logit}\,\pi_{ij} &= \beta_0 + \beta_1(r_{ij}) + s_i, \\ \text{Var}[p_{ij}] &= \phi\pi_{ij}(1-\pi_{ij})/n_{ij}, \\ s_i &\sim iid\ \text{Normal}(0, \sigma_s^2), \\ i &\in \{1, 2, \ldots, 99\}, \\ j &\in \{1, 2, 3\}, \end{aligned} \tag{4.9}$$

where p is the proportion of extreme fixations (either short or long, depending on the analysis), π is the probability being modeled, n is the number of extreme fixations, s is the subject random effect, i is the subject index, and j is the reading mode (i.e., N, P, and S) index. In my analysis, I focus on first- and second-pass reading separately.

A minimum fixation needs to be defined even for extremely short fixations analysis like this one. Here, I use fixations being 30 ms in duration or longer. Note that fixations and saccades could start to be indiscernible below that threshold.

Results and Discussion

Figures 4.15 and 4.16 (which have the same layout as Figure 4.14 described in the previous section) show the results of fitting model (4.9) to extremely short fixation data. It appears that extremely short fixations are significantly more likely to occur during bouts of probe-caught mindless reading as compared to normal reading. This pattern starts becoming evident 60 seconds prior to an onset of a thought-sampling probe and remains reliable for another 45 seconds with the p-values indicating both trends ($p < .1$) and effects ($p < .05$). This N-P difference is most pronounced in the 30-second time window ($p < .01$).

Furthermore, there are three statistically significant differences ($p < .05$) involving self-caught mindless reading (two of which involve normal reading) in the two shortest time windows. While I do not feel full confidence in interpreting these three effects due to their seemingly isolated nature, they could be associated with both (a) the subject regaining their meta-awareness and subsequently (b) reporting an attention lapse they noticed in themselves. Unfortunately, the current experimental setup weaved these two fine threads into a tight fabric and it is not possible for me to disentangle them. Nevertheless, if the high incidence of extremely short fixations is indeed a marker of the meta-cognitive transition from mindless to mindful, then future reading experiments could capitalize on it and identify these transitional moments without the need to plant self-monitoring instructions in the minds of subjects.

With all that said, the probability of an extremely short first-pass fixation is small to start with and does not exceed .04 apart from the two narrowest time windows when it more than doubles for self-caught mindless reading. Extremely short fixations are similarly infrequent in second-pass reading. Furthermore, second-pass reading does not reveal any convincing differences between normal and mindless reading.

Figures 4.17 and 4.18 show the results of fitting model (4.9) to extremely long fixation data. Extremely long fixations are clearly more likely to occur in both kinds of mindless reading ($p < .01$ for most time windows) and during both first- and second-pass reading.

Additionally, there is strong evidence that self-caught mindless reading results in significantly higher proportion of extremely long fixations than does probe-caught mindless reading. For first-pass reading, this separation is evident fairly close to the onset of a probe or a self-report (i.e., time windows of 2–10 seconds; $p < .01$). For

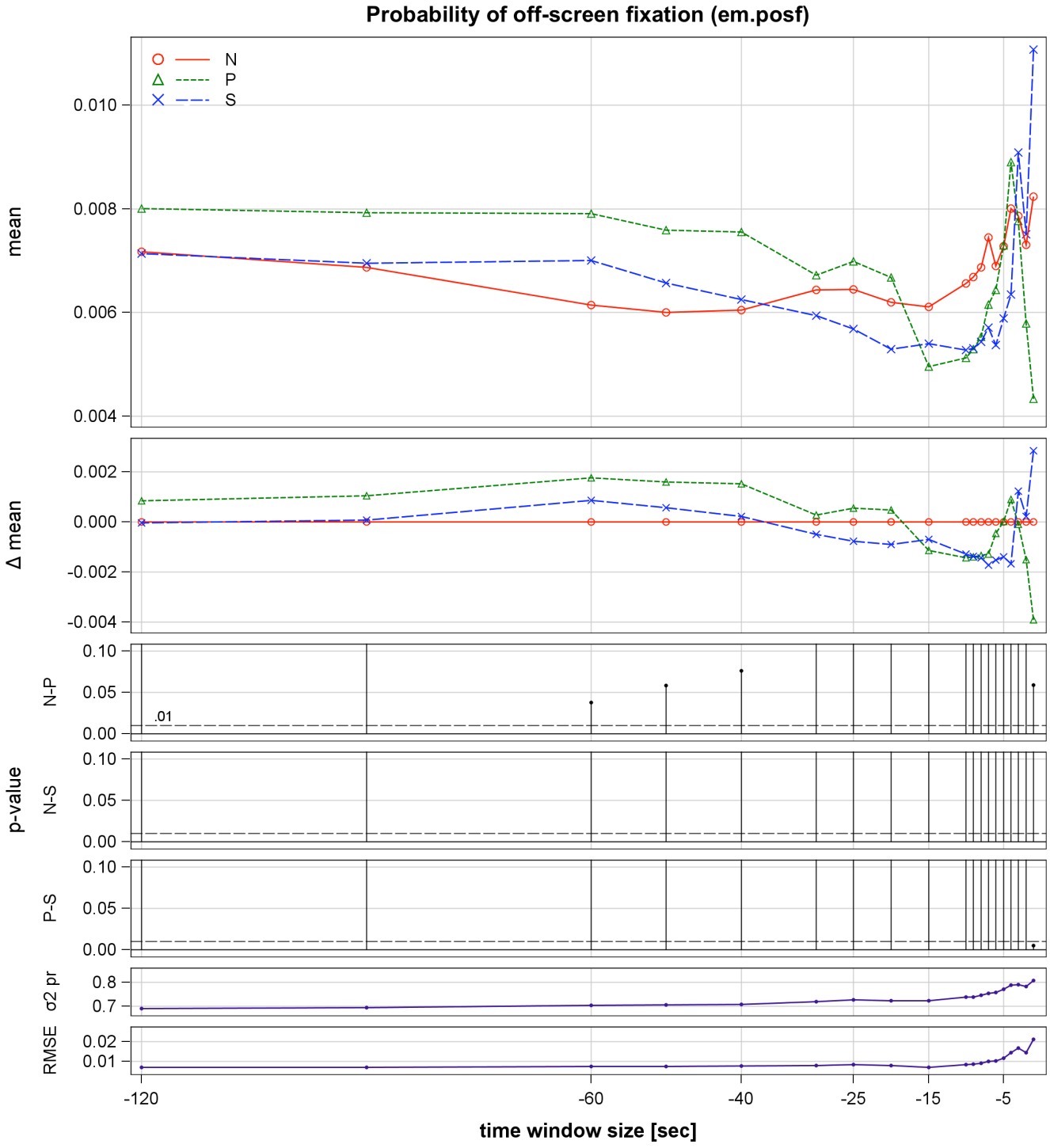


Figure 4.14: Probability of off-screen fixation ($em.posf$) for normal reading (N), probe-caught mindless reading (P), and self-caught mindless reading (S). Based on fitting model (4.8). Because the experimental setup forced below-the-screen fixations, they were excluded to avoid contamination.

second-pass reading, this pattern is visible in time windows of 7–25 seconds ($p < .05$). Similar to my earlier speculations, it is possible that meta-cognitive processes are responsible for making the eyes of a reader take those prolonged rests more frequently than would otherwise be expected if they were engaged in normal or even probe-caught mindless reading. However, unlike the case of extremely short fixations, extremely long fixations during self-caught mindless reading start becoming apparent much earlier than just 1–2 seconds prior to an attention lapse self-report. Consequently, it is plausible that they are markers of the attention beginning to wane and wax which eventually leads the reader to become aware of their mind-wandering.

Extremely long fixations appear to be up to two orders less likely to occur compared to extremely short fixations: In most time windows their probability did not exceed .0008.

Note that while fitting model (4.9) to the extremely long fixation data I was not able to achieve algorithm convergence for time windows of 1–6 seconds and as a consequence these data points are missing in Figure 4.18.

Conclusions

During both normal and mindless reading a first-pass fixation is likely to be extremely short less than four times out of a 100. Extremely long fixations on the other hand are as much as 100 times less likely than that. Therefore, despite statistically significant differences between normal and mindless reading such rare events represent limited utility for mindless reading detection. That is, while these anomalous fixations could be indicative of mindless reading, they should be used as a secondary rather than a primary indicator. Additionally, because longer episodes of mindless reading are likely to "produce" a larger number of extreme fixations, these fixations are expected to fare better in identifying long as opposed to brief attention lapses.

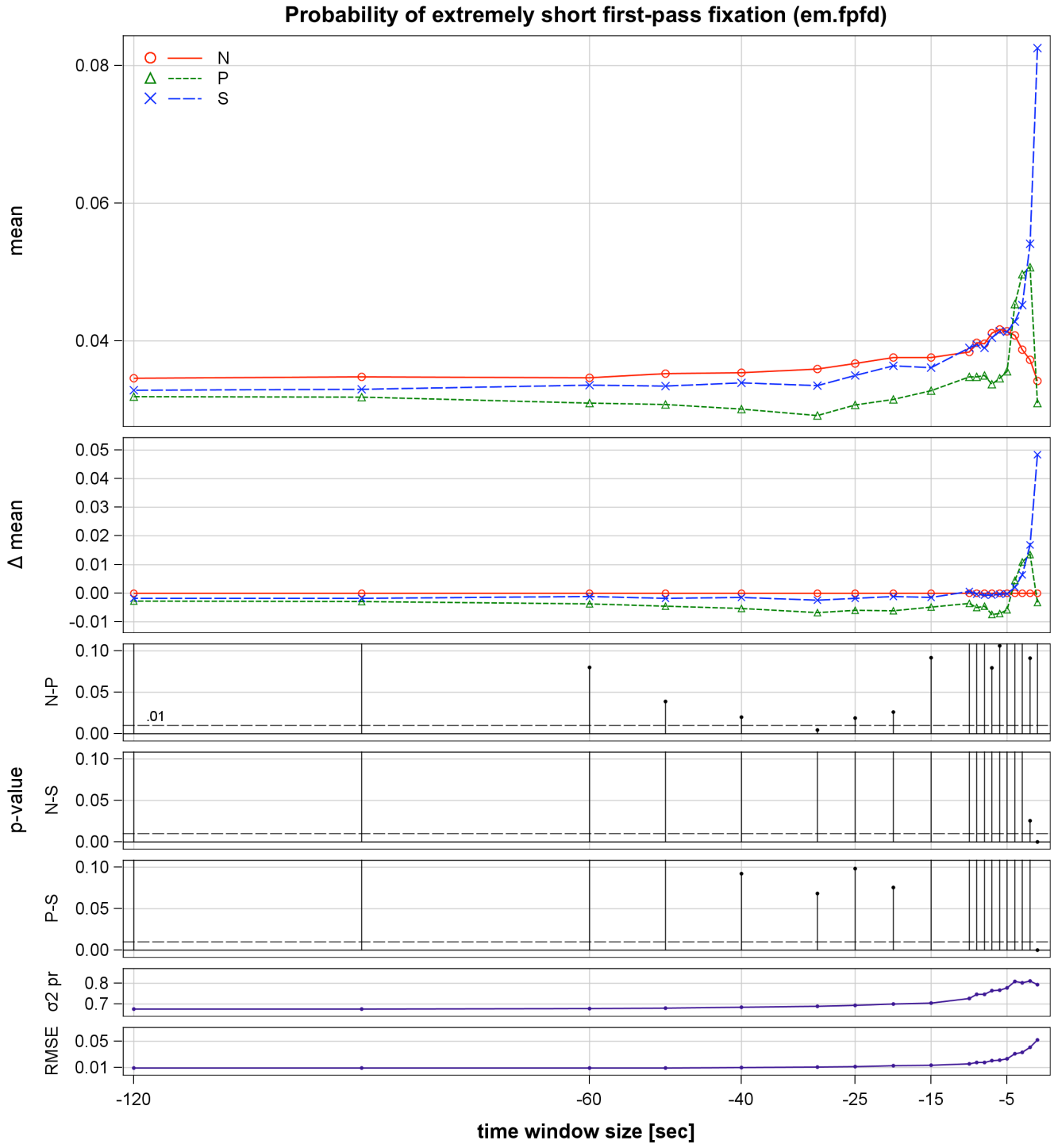


Figure 4.15: Probability of first-pass fixation duration ($em.fpfd$) shorter than 80 ms occurring in normal reading (N), probe-caught mindless reading (P), and self-caught mindless reading (S). Based on fitting model (4.9).

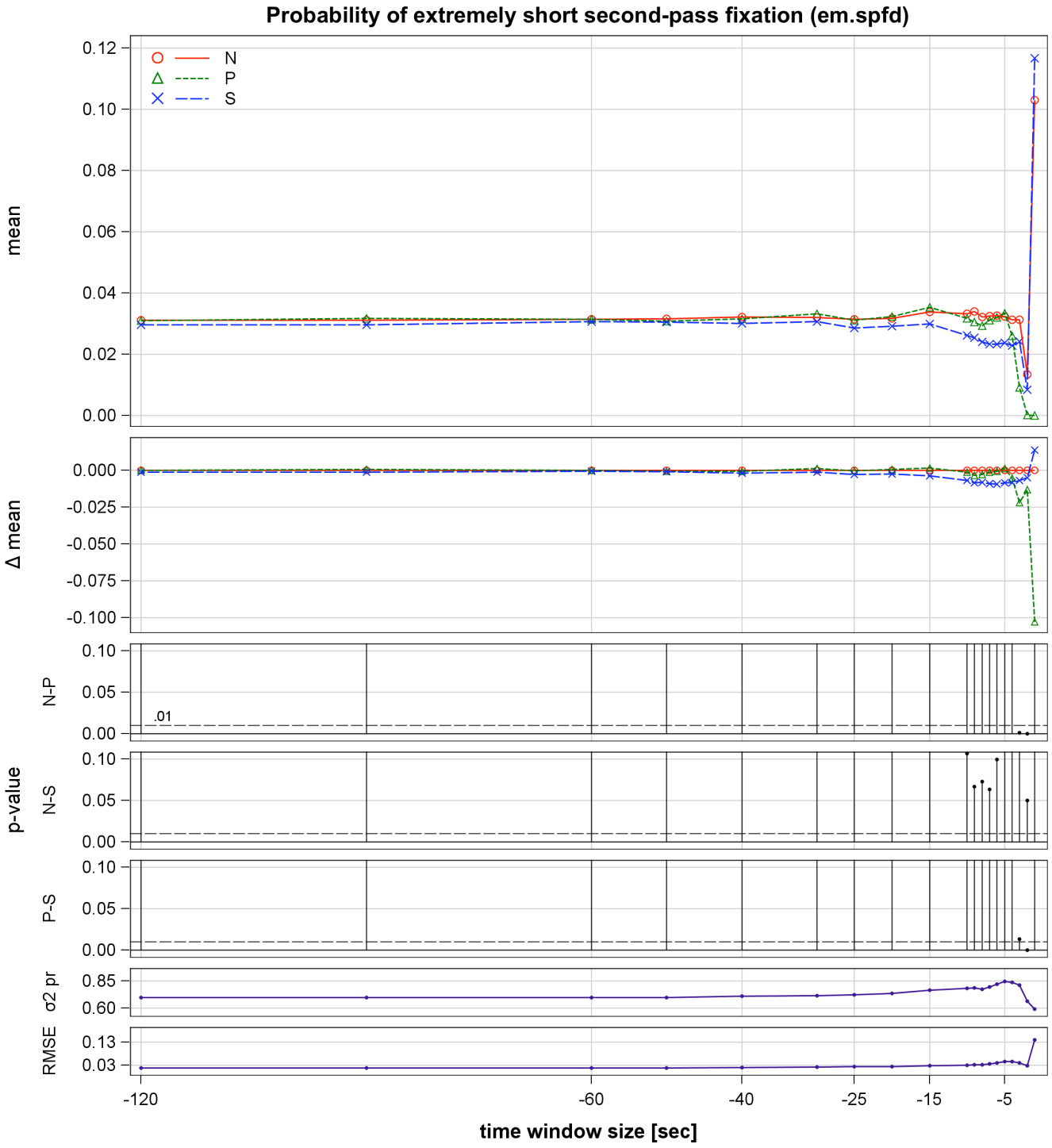


Figure 4.16: Probability of second-pass fixation duration ($em.spfd$) shorter than 80 ms occurring in normal reading (N), probe-caught mindless reading (P), and self-caught mindless reading (S). Based on fitting model (4.9).

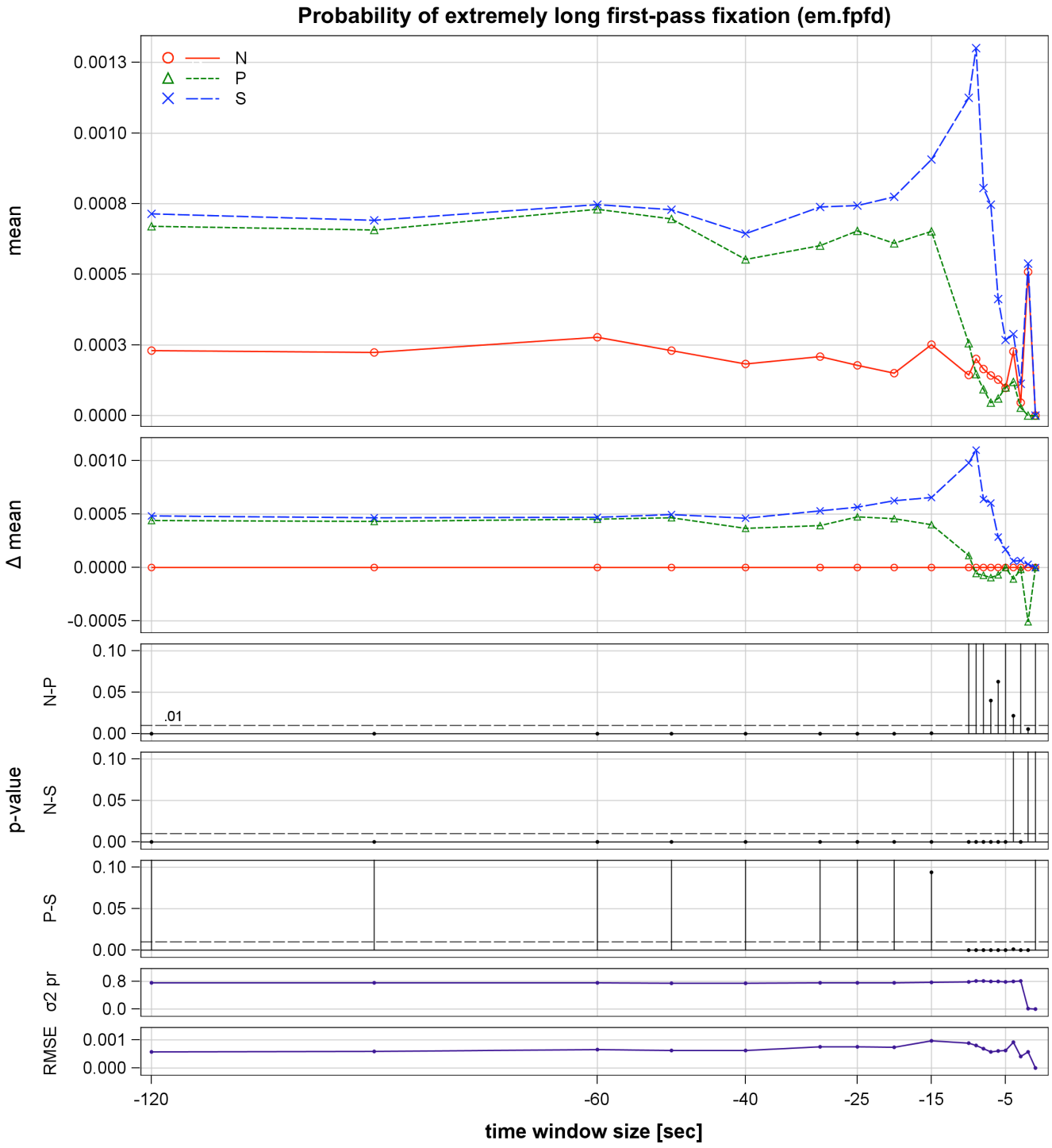


Figure 4.17: Probability of first-pass fixation duration ($em.fpfd$) longer than 1000 ms occurring in normal reading (N), probe-caught mindless reading (P), and self-caught mindless reading (S). Based on fitting model (4.9).

4.10 Lag and Successor Effects

Motivation

The extent to which lexical properties of words *n-1* and *n+1* influence fixation durations on word n (i.e., *lag* and *successor* effects, respectively) is highly contested (cf., Kliegl, Nuthmann, & Engbert, 2006; Rayner et al., 2007). One explanation of such findings and their intermittent nature is that attention is dynamically modulated during reading, occasionally encompassing multiple words and thereby giving rise to these effects. In this section, I examine the possibility that this dynamic modulation of attention could be a result of mindless reading. The results of this investigation are important because if mindless reading is indeed responsible for widening the attentional spotlight so that it includes words adjacent to the currently fixated one, then I need to account for that in my modeling endeavors. Naturally, my findings will also speak to the mentioned serial-vs-parallel word-processing debate (cf. Reichle et al., 2009; **?**).

Below, I perform two analyses. In the first one (i.e., the overall trend analysis), I examine the existence of lag and successor effects in the entire eye-movement record of my experiment. The rationale behind this is to establish if the lag and successor effects can be expected to occur during reading of extended passages of text, without dissecting it into normal and mindless reading. That it is possible that these effects are apparent in those circumstances have been suggested by Kliegl, Nuthmann, & Engbert (2006). Both that study and the current experiment examine large corpora of text. However, while Kliegl, Nuthmann, & Engbert (2006) looked at reading of individual sentences, I look at reading of long segments of real text for extended periods of time, i.e., I study an ecologically valid task.

I commence this first analysis by investigating the factors that could possibly contribute to lag and successor effects. To that effect, I look at three types of variables: (a) oculomotor variables (i.e., the length of the saccade into and out of word *n*), (b) visual variables (i.e., the lengths of words *n-1*, *n*, and *n+1*), and (c) lexical variables (i.e., the frequencies of words *n-1*, *n*, and *n+1*). It it important to note, that most of the controversy surrounding successor effects has been focused on the question of whether or not the frequency of word *n+1* influences the time spent fixating word *n* (Rayner et al., 2003). The exhaustive list of the explanatory variables I investigate is as follows (in the order in which they appear in all figures and tables):

1. Length of the previous word ($w.len^{[n-1]}$)
2. Length of the current word ($w.len^{[n]}$)
3. Length of the next word ($w.len^{[n+1]}$)
4. Frequency of the previous word ($w.frq^{[n-1]}$)
5. Frequency of the current word ($w.frq^{[n]}$)
6. Frequency of the next word ($w.frq^{[n+1]}$)
7. Length of the incoming saccade ($sacc.in.len$)
8. Landing site eccentricity ($sacc.ecc$)
9. Length of the outgoing saccade ($sacc.out.len$)

In the second analysis, I assess the influence of these variables during normal and mindless reading. That comparative analysis attempts to determine if the existence and/or the magnitude of any observed lag and successor effects are modulated by the attentive state of the reader – that is, to determine if lag and successor effects are influenced by whether the participants are reading normally versus mindlessly.

In both of these analyses, I assess the lag and successor effects using three eye-movement duration measures calculated on the current word: gaze duration ($em.fp.gd$), single fixation duration ($em.fp.sfd$), and total time ($em.oa.tt$). Both gaze duration and single fixation duration are first-pass reading measures. Gaze duration is the most important response variable because is yields the biggest statistical power. That is because while single fixation duration is defined only for words fixated exactly once, Gaze duration is defined for all words fixated at least once. In effect, single fixation duration has missing values for all words fixated more than once, while gaze duration does not. Furthermore, it is the measure that has been most closely connected to lexical processing (Rayner, 1998, 2009). That is also why, below, I consider gaze duration first. Finally, total times are important because they provide an index of the overall processing a word receives.

Model

To investigate the lag and successor effects, I set up the following statistical model

$$
\begin{aligned}
\log y_{ij} | s_i &\sim \text{Normal}(\mu_{ij}, \sigma^2_{ij}),\\
\mu_{ij} &= \beta_0\\
&\quad + \beta_1(w.len_{ij}^{[n-1]}) + \beta_2(w.len_{ij}^{[n]}) + \beta_3(w.len_{ij}^{[n+1]})\\
&\quad + \beta_4(w.frq_{ij}^{[n-1]}) + \beta_5(w.frq_{ij}^{[n]}) + \beta_6(w.frq_{ij}^{[n+1]})\\
&\quad + \beta_7(s.len.in_{ij}) + \beta_8(s.land.ecc_{ij}) + \beta_9(s.len.out_{ij})\\
&\quad + s_i,\\
s_i &\sim iid\ N(0, \sigma^2_s),\\
i &\in \{1, 2, \ldots, 99\},
\end{aligned}
\tag{4.10}
$$

where y is the fixation duration variable (i.e., $em.fp.gd$, $em.fp.sfd$, and $em.fp.tt$), s is the subject random effect, i is the subject index, and j is the word index. I used

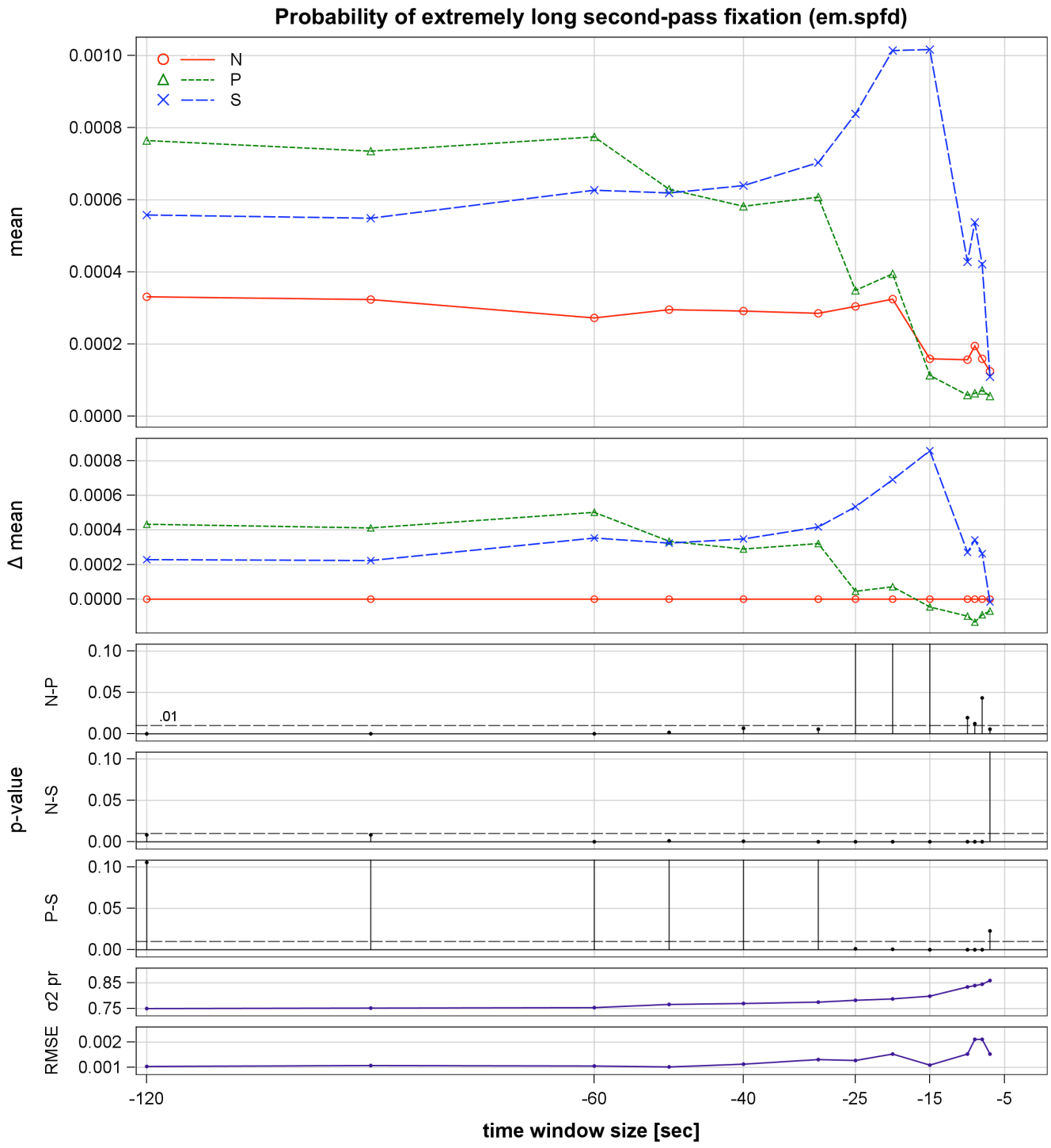


Figure 4.18: Probability of second-pass fixation duration ($em.spfd$) longer than 1000 ms occurring in normal reading (N), probe-caught mindless reading (P), and self-caught mindless reading (S). Based on fitting model (4.9). Data points for time windows of 1–6 seconds missing due to lack of convergence.

the natural log of word frequency as given by the SUBTLEX norms (Brysbaert & New, 2009) and all other variables were measured in character spaces.

As is evident from the definition of model (4.10), I assumed lognormal distribution of the residuals. As indicated by the Pearson residuals diagnostic panels (an example one for gaze duration shown in Figure 4.19[2]). Normal distribution of residuals was assumed in the right panel and lognormal was assumed in the left panel. A comparison of the two panels indicates that the lognormal distribution provides a better fit of the data while the normal distribution yields biased results. Figure 4.20 shows three examples of Pearson residuals diagnostics panels for the normal-vs-mindless comparative analysis indicating again a good model fit.

Data

Because I investigate the effect that characteristics of the previous, current, and next words have on eye movement variables on the current word, I fitted model (4.10) to data consisting of word triplets (previous-current-next) that received any number of first-pass fixations and being fixated in sequence, from left to right. More specifically, I enforced the following word-triplet selection criteria:

1. The three words are adjacent (i.e., no word skipping).
2. Each of the three words receives at least one fixation (but can receive more).
3. Each of the three words is 1-14 characters in length (0.25% of word triplets were excluded because they contained words longer than 14 characters).
4. The current word (i.e., word n) is not the first one nor the last one in line (i.e., word triplets "connected" by a return saccade are excluded).
5. Each of the three words has a frequency rating.

In the reminder of this section, I refer to the above set of rules as *the selection criteria*.

Table 4.10 shows numbers and percentages of (1) first-pass fixations made on all three words in a triplet, (2) all fixations made on all three words in a triplet, (3) single-fixation durations on the current word only, and (4) gaze durations on the current word only. Each of these four is shown for both all the word triplets and only those that met the selection criteria. The figures for total time and incoming/outgoing saccades are identical to those for gaze durations and are therefore not shown. As indicated by the second row of the table, approximately 9% of first-pass fixations on the current word were consistent with the selection criteria; these are therefore the data that the overall trend analysis is based on. As indicated further by the data in the table, the comparative analysis of normal versus mindless reading is based on 5-23% of those 9% of first-pass fixations. This final culling reflects the fact that this comparative analysis was limited to eye-movement data obtained from time intervals preceding the onset of a probe or a self-reported mind-wandering.

[2]The panels for the other dependent measures look very similar; I do not present them here for the sake of brevity.

4.10.1 Overall Trend Analysis

Gaze Durations

Figure 4.21 and Table 4.11 show the effects that the nine explanatory variables had on the gaze duration on the current word (i.e., word n). The three top subplots respectively show how the length of the previous, current, and next word affected the gaze duration on the current word. As indicated by these subplots, gaze durations increase with the length of both the current and next word ($\beta = 6.6$, $p < .0001$; $\beta = 1.2$, $p < .0001$, respectively). However, gaze durations decrease with the increase of the previous word's length ($\beta = -.17$, $p < .0001$). While the effect size for the current word is evident (about 100 ms), the lag and successor effect produced a much less pronounced effect sizes (about 20 ms).

The three middle subplots show the effects of frequencies of the previous, current, and next words on the current word's gaze durations. The effect of the current word's frequency is the only significant effect here ($\beta = -3.5$, $p < .0001$): Gaze duration on the current word increases as its frequency decreases. Neither the lag nor successor word-frequency effects are significant ($p \geq .2625$).

Finally, the three bottom subplots show how the incoming saccade length, fixation landing-site eccentricity on the current word, and outgoing saccade length affect the gaze durations on the current word. The effects of both saccade lengths are significant. As expected, longer incoming saccades are associated with less parafoveal processing of the next word and thus result in longer gaze durations on the current word ($\beta = 3.4$, $p < .0001$). Furthermore, longer outgoing saccades yield shorter gaze durations on the current word ($\beta = -2.2$, $p < .0001$). Finally, the slope for the landing site eccentricity was large, positive, and statistically reliable ($\beta = 4.4$, $p < .0001$), indicating that the shortest gaze durations are associated with the initial fixations located near the center or to the left of center of the current word. Note that the *preferred viewing position* is to the left of word center (McConkie et al., 1988) and therefore these fixations are likely responsible for left-of-center gaze durations being the shortest.

Single Fixation Durations

Figure 4.22 and Table 4.12 show the effects that the nine explanatory variables had on the single-fixation durations

Table 4.10: Numbers and percentages of: (1) first-pass fixations on word triplets, (2) all fixations (i.e., both first- and second-pass), (3) single fixation durations (SFDs; $em.fp.sfd$) on the current word only, and (4) gaze durations (GDs; $em.fp.gd$) on the current word only. Each of these four is shown for two classes of word triplets: All those fixated in sequence (Type = A) and those that met the selection criteria (Type = C). t denotes the time-window size counted back from an onset of a probe or a self-reported mind-wandering (t_0). The first two rows (i.e., those for $t = 0$) show data when no time window interval was enforced (i.e., ignoring all probe and self-reported mind-wandering events) and therefore summarize data used in the overall trend analysis. The subsequent rows summarize data for the normal-vs-mindless reading comparative analysis. The figures for gaze durations (i.e., the last 3 columns) naturally coincide with those for incoming/outgoing saccades. Missing values due to incomplete frequency ratings are reflected in the table.

t	type	First-pass fix. (word triplet)			All fix. (word triplet)			SFD (current word)			GD (current word)		
		N	%A[1)]	%C[2)]	N	%A	%C	N	%A	%C	N	%A	%C
0	A	2 133 371	100.00	–	3 307 432	100.00	–	636 780	100.00	–	758 860	100.00	–
	C	**192 691**	**9.03**	**100.00**	**270 845**	**8.19**	**100.00**	**45 072**	**7.08**	**100.00**	**52 909**	**6.97**	**100.00**
10	A	113 100	100.00	–	152 628	100.00	–	33 747	100.00	–	40 263	100.00	–
	C	**10 176**	**9.00**	**5.28**	**12 491**	**8.18**	**4.61**	**2 339**	**5.81**	**5.19**	**2 779**	**6.90**	**5.25**
15	A	162 968	100.00	–	229 218	100.00	–	47 990	100.00	–	57 513	100.00	–
	C	15 379	9.44	7.98	19 566	8.54	7.22	3 537	6.15	7.85	4 193	7.29	7.92
20	A	207 973	100.00	–	300 763	100.00	–	61 142	100.00	–	73 318	100.00	–
	C	20 075	9.65	10.42	26 223	8.72	9.68	4 617	6.30	10.24	5 472	7.46	10.34
25	A	250 837	100.00	–	367 933	100.00	–	73 711	100.00	–	88 379	100.00	–
	C	24 657	9.83	12.80	32 425	8.81	11.97	5 686	6.43	12.62	6 721	7.60	12.70
30	A	289 661	100.00	–	429 702	100.00	–	84 829	100.00	–	101 957	100.00	–
	C	**28 545**	**9.85**	**14.81**	**37 952**	**8.83**	**14.01**	**6 509**	**6.38**	**14.44**	**7 744**	**7.60**	**14.64**
40	A	357 040	100.00	–	537 857	100.00	–	104 449	100.00	–	125 691	100.00	–
	C	35 451	9.93	18.40	47 896	8.90	17.68	8 097	6.44	17.96	9 613	7.65	18.17
50	A	411 152	100.00	–	626 956	100.00	–	120 461	100.00	–	145 117	100.00	–
	C	40 881	9.94	21.22	55 429	8.84	20.47	9 336	6.43	20.71	11 071	7.63	20.92
60	A	452 106	100.00	–	698 654	100.00	–	132 712	100.00	–	160 072	100.00	–
	C	45 157	9.99	23.43	61 899	8.86	22.85	10 261	6.41	22.77	12 206	7.63	23.07

[1)] percentage of all data with the same t (the row just above)

[2)] percentage of all data that met the selection criteria (i.e., those with t=0 and type=C; the second row)

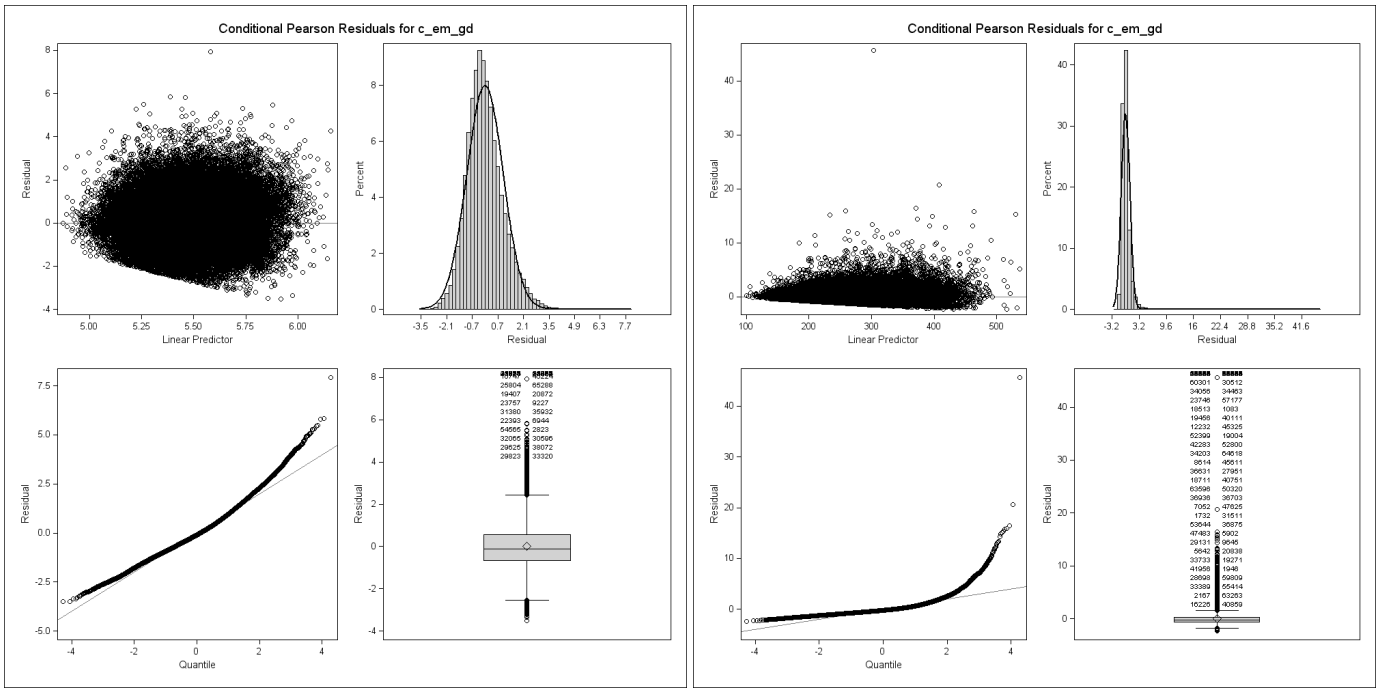


Figure 4.19: Conditional Pearson residuals diagnostic panels for model 4.10 with gaze duration ($em.fp.gd$) on the current word as the response variable fitted to the overall trend data partition. The left plot is a result of assuming the lognormal distribution of residuals, while the right one of assuming the normal distribution. Residuals panels for the other duration measures and both the overall trend and mind-wandering data partitions look similar. The left panel shows good fit, while the right one a clear misspecification. See Figure 4.20 for an example of Pearson residuals diagnostics panels for the normal-vs-mindless comparative analysis.

on the current word. As the top subplots show, only the lag length effect is significant ($\beta = -0.7$, $p = .0016$), indicating that longer previous words are negatively correlated with shorter single-fixation durations on the current word.

The middle subplots show that higher frequencies of both the previous ($\beta = -0.9$, $p = .0383$) and current ($\beta = -1.9, p < .0001$) words are associated with a decrease in single-fixation durations on the current word with he lag effect being approximately half the size of the immediacy effect.

Finally, as shown in the bottom subplots, similarly to gaze durations, single-fixation duration on the current word increases as the incoming saccade length increases ($\beta = 3.1$, $p < .0001$) and as the outgoing saccade length decreases ($\beta = -0.6, p < .0001$). Additionally, the effect of landing-site eccentricity is significant as well ($\beta = -1.7$, $p < .0001$) but appears more erratic when compared to gaze durations.

Total Times

Figure 4.23 and Table 4.13 show the effects of the nine explanatory variables on the total times on the current word. The pattern of effects is exactly that same as for gaze durations, but with effects being larger in magnitude. Evidently, adding second-pass reading fixations does not change the results much. Due to the similarities between the total-time and gaze-duration results I do not discuss the former any further here.

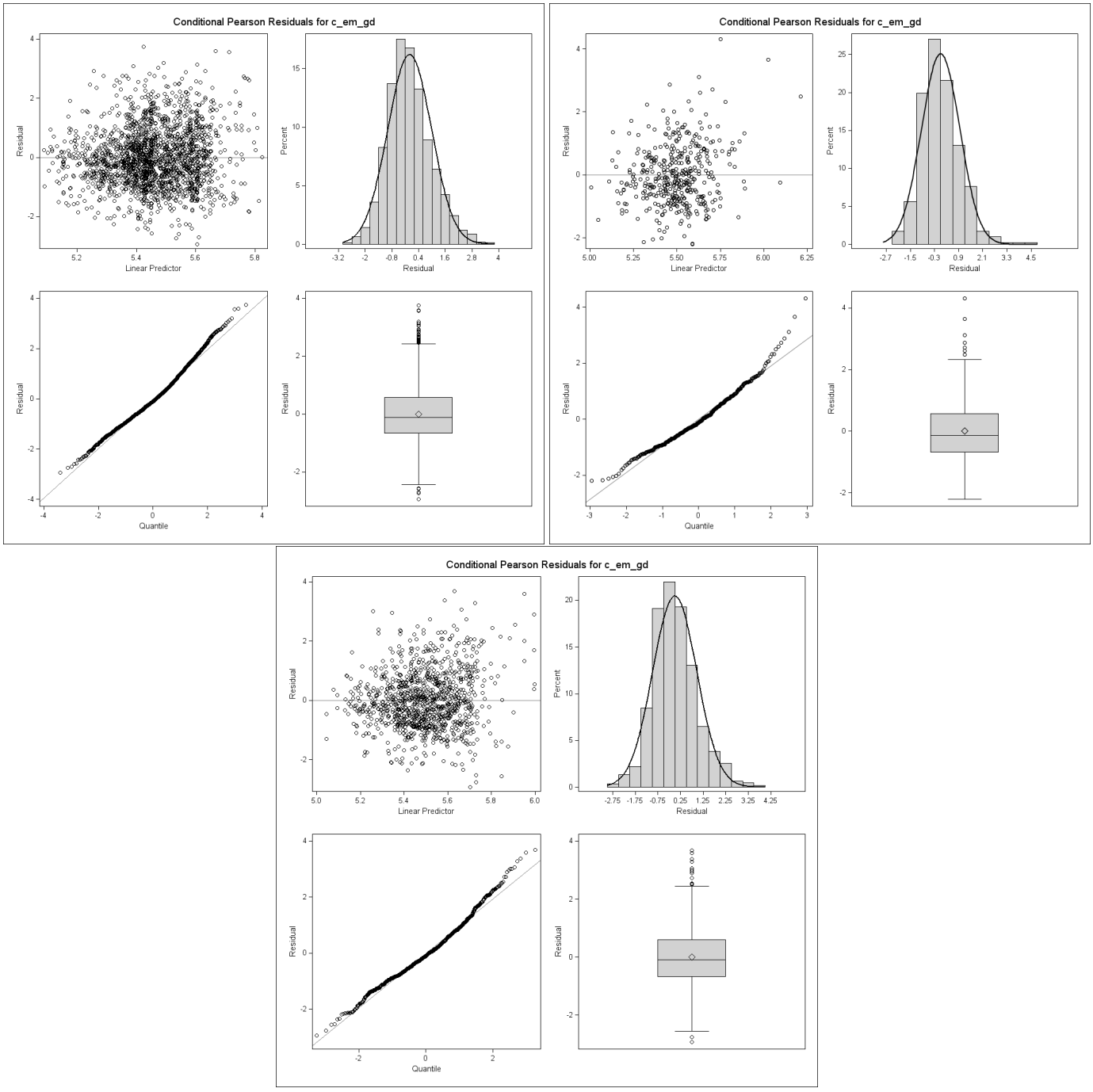


Figure 4.20: Conditional Pearson residuals diagnostic panels for model 4.10 with gaze duration ($em.fp.gd$) on the current word as the response variable fitted to the 10-second time window of the normal reading (top left), probe-caught mindless reading (top right), and self-caught mindless reading (bottom) data. Residuals panels for the other duration measures and time windows look similar. See Figure 4.19 for an example of residual diagnostics panels for the overall trend analysis.

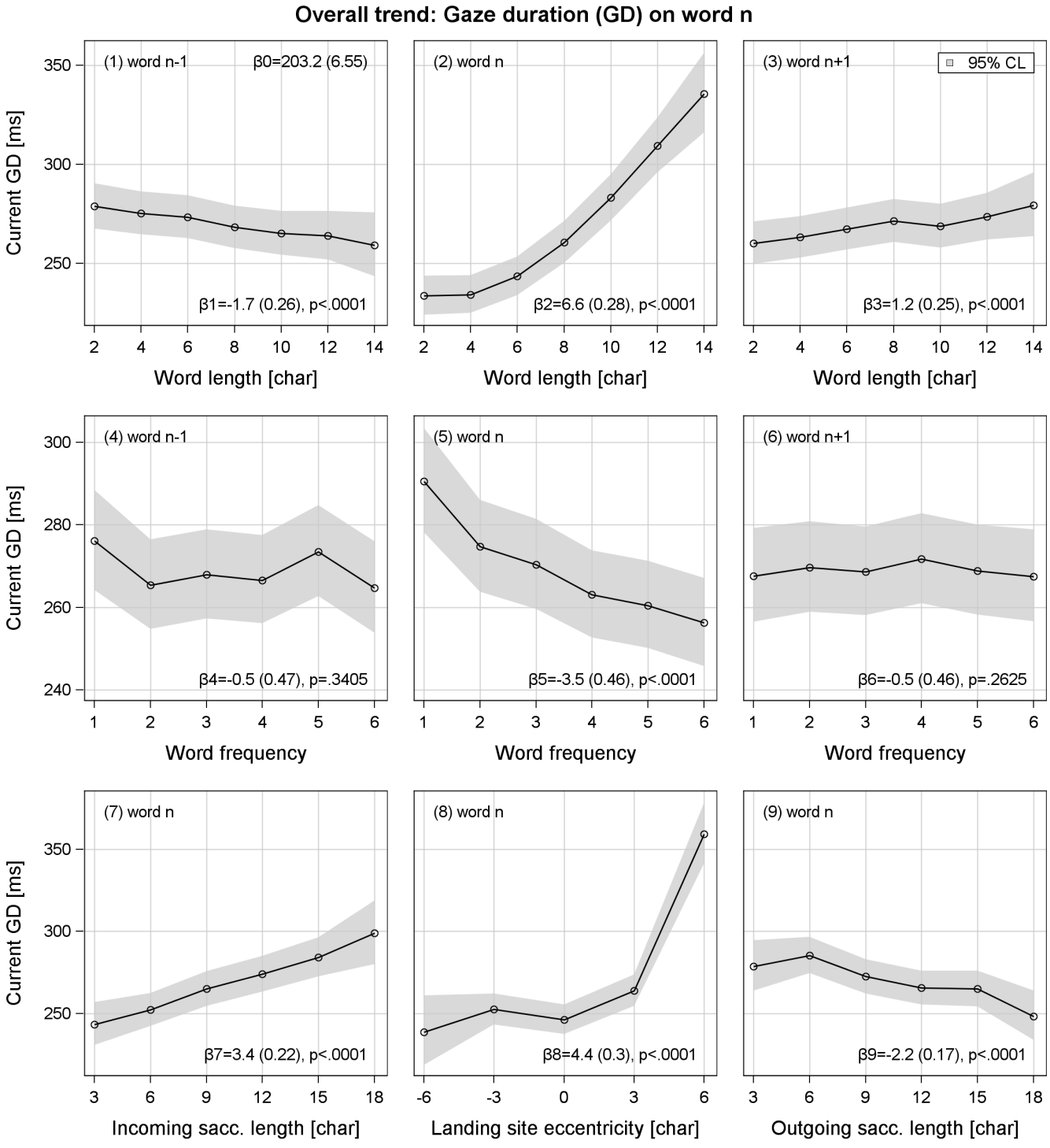


Figure 4.21: Effects of word length (in character spaces), word frequency (natural log; SUBTLEX, Brysbaert & New, 2009), length of an incoming and outgoing saccade (in character spaces), and landing site eccentricity (in character spaces) of the previous (n-1), current (n), and next (n+1) word on gaze duration ($em.fp.gd$) on the current word (model (4.10)). Values plotted are predicted population margins (least square means). Data shown in Table 4.11.

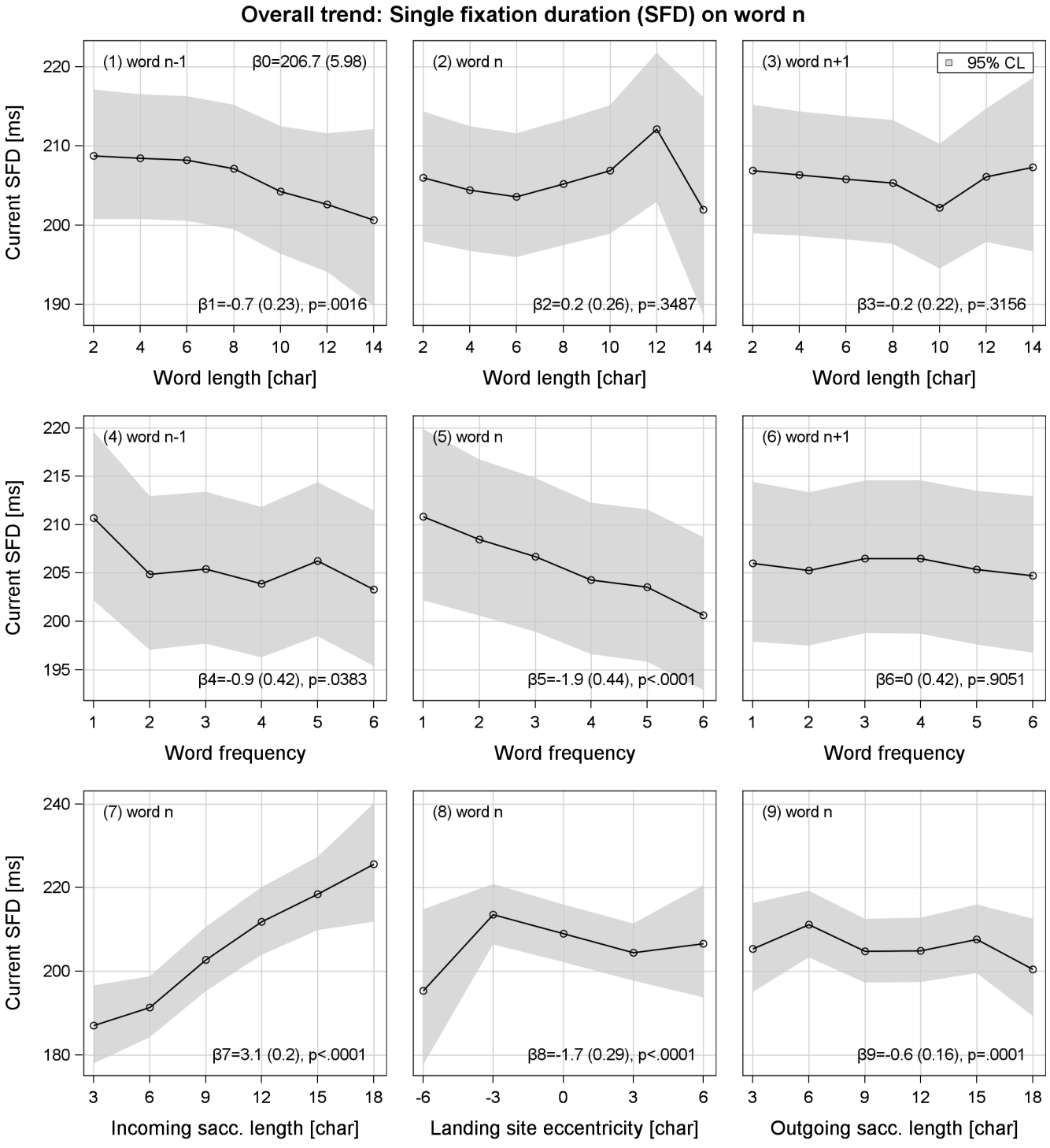


Figure 4.22: Effects of word length (in character spaces), word frequency (natural log; SUBTLEX, Brysbaert & New, 2009), length of an incoming and outgoing saccade (in character spaces), and landing site eccentricity (in character spaces) of the previous (n-1), current (n), and next (n+1) word on single fixation duration ($em.fp.sfd$) on the current word (model (4.10)). Values plotted are predicted population margins (least square means). Data shown in Table 4.12.

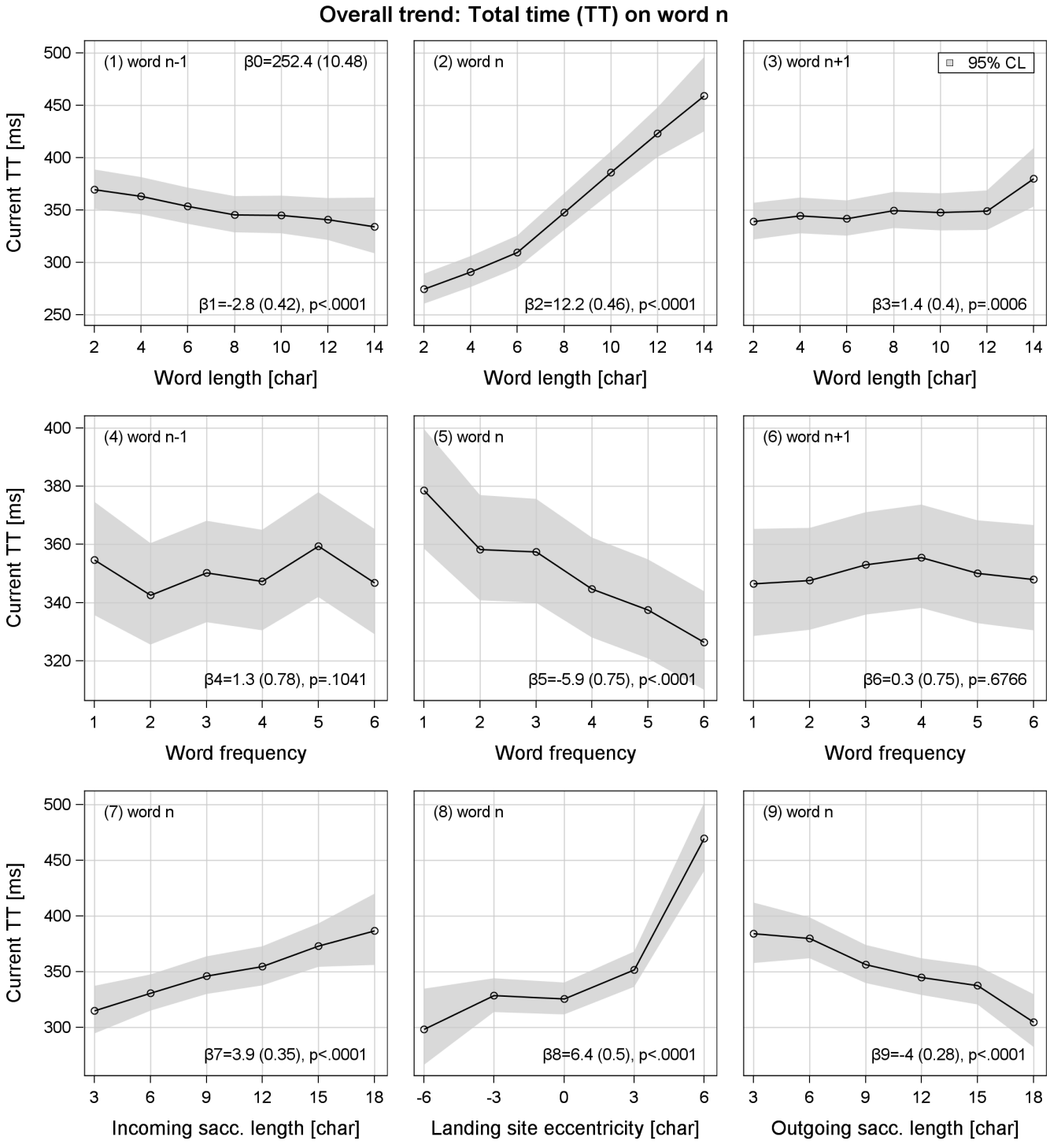


Figure 4.23: Effects of word length (in character spaces), word frequency (natural log; SUBTLEX, Brysbaert & New, 2009), length of an incoming and outgoing saccade (in character spaces), and landing site eccentricity (in character spaces) of the previous (n-1), current (n), and next (n+1) word on total time (*em.oa.tt*) on the current word (model (4.10)). Values plotted are predicted population margins (least square means). Data shown in Table 4.13.

Table 4.11: The first four rows show the regression coefficients, theirs standard errors, and the corresponding p-values for the effects of (1) word length (in character spaces), (2) word frequency (natural log; SUBTLEX, Brysbaert & New, 2009), (3) length of an (4) incoming and (5) outgoing saccade (in character spaces), and (6) landing site eccentricity (in character spaces) of the previous (n-1), current (n), and next (n+1) word on gaze duration ($em.fp.gd$) on the current word (model (4.10)). The subsequent rows show predicted population margins (least square means; MSE in brackets) of the gaze duration ($em.fp.gd$) on the current word. Data plotted in Figure 4.21.

x	Word length ($w.len$)						Word frequency ($w.frq$)						Saccade ($sacc$)					
	n-1		n		n+1		n-1		n		n+1		$len.in$		$land.ecc$		$len.out$	
β_0	203.16 (6.55)																	
β_i	-1.68		6.57		1.16		-0.45		-3.49		-0.52		3.41		4.39		-2.15	
β_{SE}	0.26		0.28		0.25		0.47		0.46		0.46		0.22		0.3		0.17	
p	$<.0001$		$<.0001$		$<.0001$		0.3405		$<.0001$		0.2625		$<.0001$		$<.0001$		$<.0001$	
-6															239	(11.0)		
-3															253	(4.9)		
0															246	(4.6)		
1							276	(6.3)	291	(6.5)	268	(5.9)						
2	279	(5.9)	234	(5.1)	260	(5.5)	265	(5.6)	275	(5.7)	270	(5.6)						
3							268	(5.5)	270	(5.6)	269	(5.5)	243	(6.7)	264	(5.0)	279	(8.0)
4	275	(5.6)	234	(4.9)	263	(5.4)	267	(5.5)	263	(5.4)	272	(5.6)						
5							274	(5.7)	260	(5.4)	269	(5.6)						
6	273	(5.6)	243	(5.1)	267	(5.5)	265	(5.7)	256	(5.5)	267	(5.7)	252	(5.2)	359	(9.5)	285	(5.7)
7																		
8	268	(5.5)	261	(5.4)	271	(5.6)												
9													265	(5.4)			272	(5.4)
10	265	(5.7)	283	(6.0)	269	(5.7)												
11																		
12	264	(6.3)	309	(7.1)	273	(6.1)							274	(5.6)			266	(5.3)
13																		
14	259	(8.3)	336	(10.4)	279	(8.3)												
15													284	(6.1)			265	(5.6)
18													299	(10.0)			248	(7.8)

Table 4.12: The first four rows show the regression coefficients, theirs standard errors, and the corresponding p-values for the effects of (1) word length (in character spaces), (2) word frequency (natural log; SUBTLEX, Brysbaert & New, 2009), (3) length of an (4) incoming and (5) outgoing saccade (in character spaces), and (6) landing site eccentricity (in character spaces) of the previous (n-1), current (n), and next (n+1) word on single fixation duration ($em.fp.sfd$) on the current word (model (4.10)). The subsequent rows show predicted population margins (least square means; MSE in brackets) of the single fixation duration ($em.fp.sfd$) on the current word. Data plotted in Figure 4.22.

x	Word length ($w.len$)						Word frequency ($w.frq$)						Saccade ($sacc$)					
	n-1		n		n+1		n-1		n		n+1		$len.in$		$land.ecc$		$len.out$	
β_0	206.73 (5.98)																	
β_i	-0.73		0.24		-0.22		-0.88		-1.91		-0.05		3.08		-1.73		-0.61	
β_{SE}	0.23		0.26		0.22		0.42		0.44		0.42		0.2		0.29		0.16	
p	0.0016		0.3487		0.3156		0.0383		$<.0001$		0.9051		$<.0001$		$<.0001$		0.0001	
-6															195	(9.7)		
-3															214	(3.7)		
0															209	(3.5)		
1							211	(4.5)	211	(4.6)	206	(4.3)						
2	209	(4.2)	206	(4.2)	207	(4.2)	205	(4.1)	209	(4.2)	205	(4.1)						
3							205	(4.0)	207	(4.1)	207	(4.1)	187	(4.8)	204	(3.5)	205	(5.5)
4	208	(4.1)	204	(4.1)	206	(4.1)	204	(4.0)	204	(4.0)	206	(4.1)						
5							206	(4.1)	204	(4.1)	205	(4.1)						
6	208	(4.1)	204	(4.0)	206	(4.0)	203	(4.2)	201	(4.1)	205	(4.2)	191	(3.8)	207	(6.9)	211	(4.1)
7																		
8	207	(4.1)	205	(4.1)	205	(4.0)												
9													203	(4.0)			205	(3.9)
10	204	(4.2)	207	(4.2)	202	(4.1)												
11																		
12	203	(4.5)	212	(4.9)	206	(4.3)							212	(4.2)			205	(3.9)
13																		
14	201	(5.8)	202	(7.1)	207	(5.7)												
15													218	(4.5)			208	(4.2)
18													226	(7.3)			201	(6.0)

Table 4.13: The first four rows show the regression coefficients, theirs standard errors, and the corresponding p-values for the effects of (1) word length (in character spaces), (2) word frequency (natural log; SUBTLEX, Brysbaert & New, 2009), (3) length of an (4) incoming and (5) outgoing saccade (in character spaces), and (6) landing site eccentricity (in character spaces) of the previous (n-1), current (n), and next (n+1) word on total time ($em.oa.tt$) on the current word (model (4.10)). The subsequent rows show predicted population margins (least square means; MSE in brackets) of the total time ($em.oa.tt$) on the current word. Data plotted in Figure 4.23.

x	Word length ($w.len$)						Word frequency ($w.frq$)						Saccade ($sacc$)					
	n-1		n		n+1		n-1		n		n+1		$len.in$		$land.ecc$		$len.out$	
β_0	252.4 (10.48)																	
β_i	-2.79		12.22		1.37		1.26		-5.94		0.31		3.94		6.36		-3.97	
β_{SE}	0.42		0.46		0.4		0.78		0.75		0.75		0.35		0.5		0.28	
p	<.0001		<.0001		0.0006		0.1041		<.0001		0.6766		<.0001		<.0001		<.0001	
-6															298	(18.0)		
-3															328	(7.8)		
0															325	(7.5)		
1							355	(10.1)	379	(10.6)	346	(9.5)						
2	370	(9.7)	274	(7.5)	339	(9.0)	343	(9.0)	358	(9.3)	348	(9.0)						
3							350	(9.0)	357	(9.2)	353	(9.1)	315	(11.1)	352	(8.2)	384	(14.1)
4	363	(9.2)	291	(7.6)	345	(8.8)	347	(8.9)	345	(8.9)	355	(9.2)						
5							359	(9.3)	337	(8.8)	350	(9.1)						
6	354	(8.9)	310	(8.0)	342	(8.7)	347	(9.3)	326	(8.7)	348	(9.3)	331	(8.5)	470	(15.9)	380	(9.5)
7																		
8	346	(8.8)	348	(9.0)	350	(8.9)												
9													346	(8.7)			356	(8.8)
10	345	(9.3)	386	(10.2)	348	(9.2)												
11																		
12	341	(10.3)	423	(12.3)	349	(9.8)							355	(9.0)			345	(8.5)
13																		
14	334	(13.9)	459	(18.4)	380	(14.5)												
15													373	(10.1)			337	(8.9)
18													387	(16.8)			305	(12.3)

4.10.2 Normal-Mindless Comparative Analysis

The overall trend analysis I have discussed above was based on fixations made during both normal and mindless reading. In this section I separate these two modes of reading into normal reading, probe-caught mindless reading, and self-caught mindless reading. Much like I have done in earlier sections of this chapter, I sample fixations indigenous to these tree conditions through a set of time windows of different size. Like before, these time windows encompass fixations which occurred prior to an onset of a thought-sampling probe or a self-report of mindless reading. Because this current analysis involved a more complicated model, I do not use the very narrow time windows I have used before and instead focus on the 10–60s range[3].

To facilitate exposition of the results, I dichotomized the nine explanatory variables as follows: (1) word length (≤ 4 vs. ≥ 5 characters), (2) landing-site eccentricity (≤ 0 vs. > 0 characters), and (3) all other variables (e.g., word frequency) were split on their respective medians. By using binary covariates I am able to focus on slopes (i.e., regression coefficients) and thus convey both the direction and magnitude of an effect in an easy-to-interpret manner. For example, a 10-ms slope for the word-length lag effect on gaze duration indicates that long (i.e., ≥ 5 characters) previous words are associated with 10-ms longer gaze durations on the current word than are short (i.e., ≤ 4 characters) previous words.

Gaze Durations

Figure 4.24 and Table 4.14 show the effects that the nine explanatory variables have on gaze durations on the current word during 10–60s of normal reading, probe-caught mindless reading, and self-caught mindless reading. Identically to the overall trend figures, the top row of subplots shows word-length effects, the middle row shows word-frequency effects, and the bottom row shows the saccade effects. The top portion of each subplot shows slopes for the associated explanatory variable and the bottom portion shows the corresponding p-values (which indicate whether or not a slope is significantly different from zero).

The top row of Figure 4.24 shows the word-length effects on gaze durations on the current word. As shown in the left subplot, none of the previous word length slopes is significant ($p > .05$). As is evident in the middle panel, the length of the current word modulated gaze durations on that word irrespective of whether the readers were reading normally or mindlessly. Interestingly, word-length effect is noticeably more pronounced during mindless reading (both probe- and self-caught) relative to normal reading. This suggests that perceptual variables play more important role in guiding the eyes of a reader when they are mind-wandering. The right panel shows weak evidence for word-length successor effects during probe-caught mindless reading.

The middle row of Figure 4.24 shows the word-frequency effects on gaze durations on the current word. The left panel indicates the presence of word-frequency lag effects during intervals of both normal and self-caught mindless reading. The direction of this effect, however, is inconsistent with the so-called *spillover effects* (Rayner & Duffy, 1986). As expected, the middle panel shows a negative correlation between the current's word frequency and gaze duration on that word during normal reading. That effect is also present during self-caught mindless reading, but its onset is delayed and magnitude attenuated relative to normal reading. As indicated further by the middle plot, prior to a reader being caught on being mindless the word-frequency immediacy effect is nowhere to be found. Overall, these results provide a partial replication of the finding that the effects of lexical variables on gaze durations are weakened or eliminated during mindless reading (Reichle, Reineberg, & Schooler, 2010). As shown in the right panel of the middle row, there is no evidence for word-frequency successor effects in either normal or mindless reading.

Finally, the bottom row of Figure 4.24 shows the saccade effects on gaze durations on the current word. As indicated by the left subplot, processing time on a word is increased after longer incoming saccades in both normal and mindless reading. This illustrates the benefits of parafoveal preview in shortening word processing times. The right subplot shows an inverse relationship between gaze duration and outgoing saccade length. This effect is clearly present during normal reading and, to a lesser degree, during probe-caught mindless reading (and yields large slopes). It is also visible during self-caught mindless reading, but the slopes are small (at most 5 ms). The middle plot contains evidence that left-of-center landing sites are associated with shorter word processing times. This likely is a manifestation of the preferred viewing position effect (McConkie et al., 1988). However, this effect is only present during normal and self-caught mindless reading, and is absent during probe-caught mindless reading. This suggests that instances of profound mind-wandering are associated with the lack of the preferred viewing position effect.

Single Fixation Durations

Figure 4.25 and Table 4.15 (which have the same layout as the gaze-duration ones) show the effects that the nine explanatory variables had on the single-fixation durations during 10–60s of normal reading, probe-caught mindless reading, and self-caught mindless reading. A comparison of Figures 4.24 and 4.25 reveals that the slopes for single-

[3] However, in data tables I do report results for time windows of 4–60s.

fixation durations are much less stable relative to the ones for gaze duration. This increased variability reflects the fact that there were considerably fewer observations with which to calculate the slopes for single-fixations (see Table 4.15).

The top row of Figure 4.25 shows the word-length effects on single-fixation durations on the current word. There is essentially no evidence for length effects across all three subplots. The situation is similar for the middle row which shows the word-frequency effects on the current word. The only exceptions to that being (a) a case of weak evidence for a anti-spillover effect during self-caught mindless reading and (b) an anticipated inverse association between the frequency of the current word and single-fixation duration on that word during normal reading. Finally, the bottom row shows the saccade effects on single-fixation durations on the current word. Just like with gaze durations, longer incoming saccades are associated with longer single fixations. Additionally, longer outgoing saccades are associated with shorter single fixations, again, mostly during normal reading and probe-caught mindless reading. Furthermore, there is little evidence for the preferred viewing position in single-fixation duration data.

Total Times

Figure 4.26 and Table 4.16 show the effects that the nine explanatory variables had on the total times on the current word during 10–60s of normal reading, probe-caught mindless reading, and self-caught mindless reading. When compared to gaze duration results I have discussed above, total times show almost identical patterns of the word-length, word-frequency, and saccade effects. Because of these similarities, I focus only on the two obvious differences between gaze durations and total times. First, all effects are amplified for total times, sometimes being twice the magnitude of gaze durations. This, together with the lack of big differences between gaze durations and total times indicates that second-pass reading proceeds in a largely similar fashion to first-pass reading. Second, it appears that the length of the previous word is negatively related to the total time on the current word during both normal and mindless reading, especially in wider time windows.

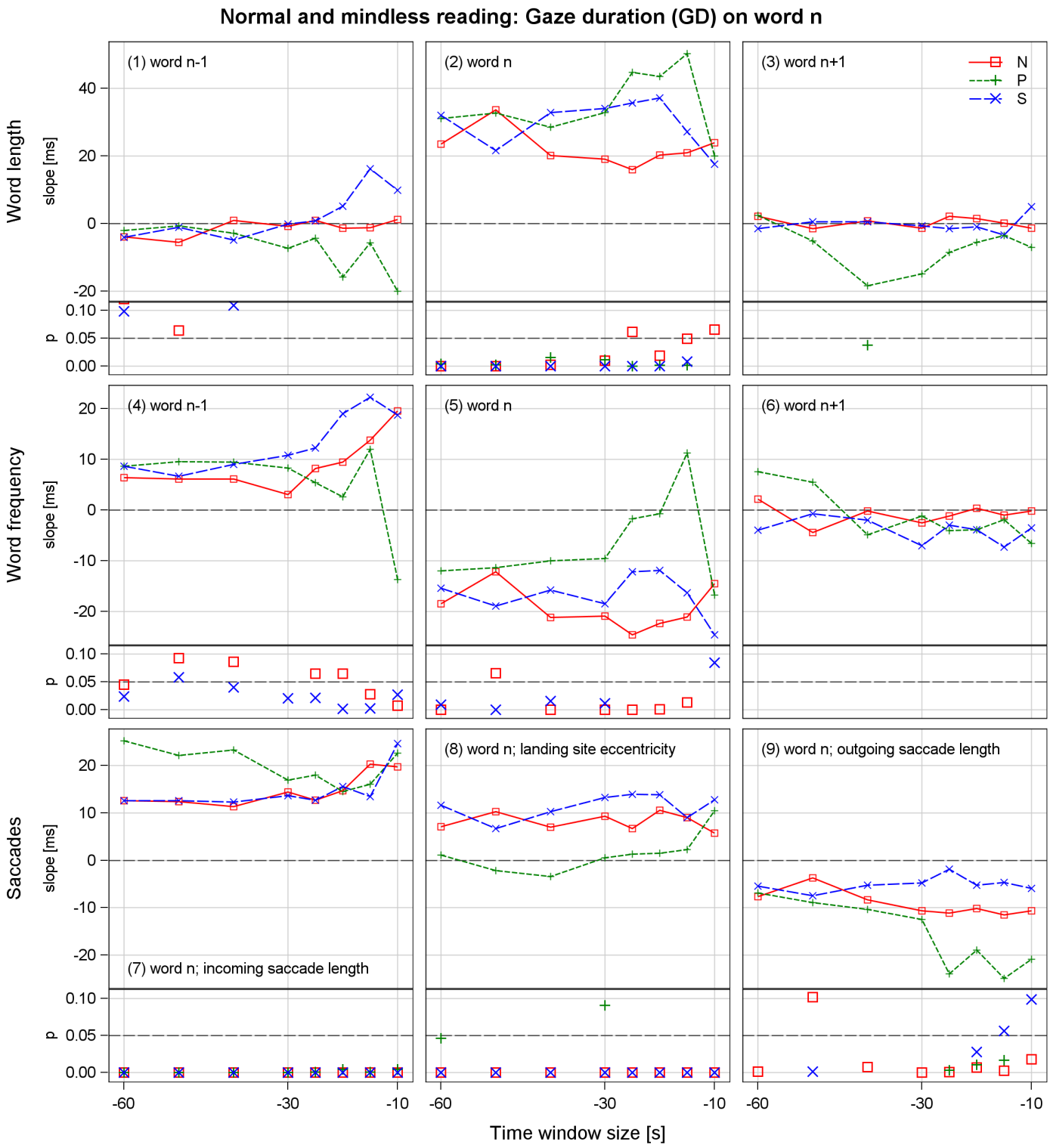


Figure 4.24: Effects (slopes) of word length (in character spaces), word frequency (natural log; SUBTLEX, Brysbaert & New, 2009), length of an incoming and outgoing saccade (in character spaces), and landing site eccentricity (in character spaces) of the previous (n-1), current (n), and next (n+1) word on gaze duration ($em.fp.gd$) on the current word for normal reading (N), probe-caught mindless reading (P), and self-caught mindless reading (S) and for time windows of 10–60s (model (4.10)). Data shown in Table 4.14.

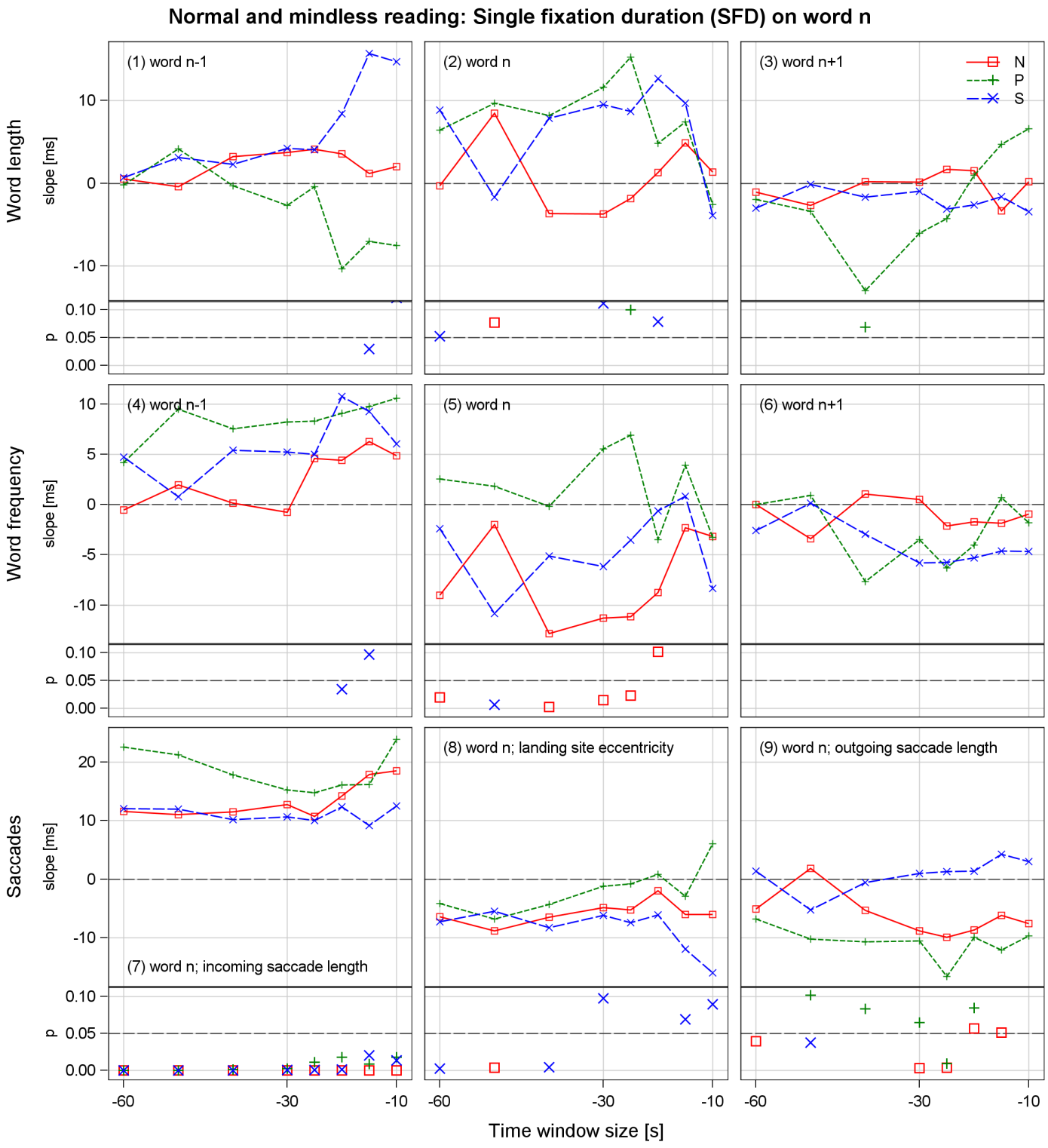


Figure 4.25: Effects (slopes) of word length (in character spaces), word frequency (natural log; SUBTLEX, Brysbaert & New, 2009), length of an incoming and outgoing saccade (in character spaces), and landing site eccentricity (in character spaces) of the previous (n-1), current (n), and next (n+1) word on single fixation duration ($em.fp.sfd$) on the current word for normal reading (N), probe-caught mindless reading (P), and self-caught mindless reading (S) and for time windows of 10–60s (model (4.10)). Data shown in Table 4.15.

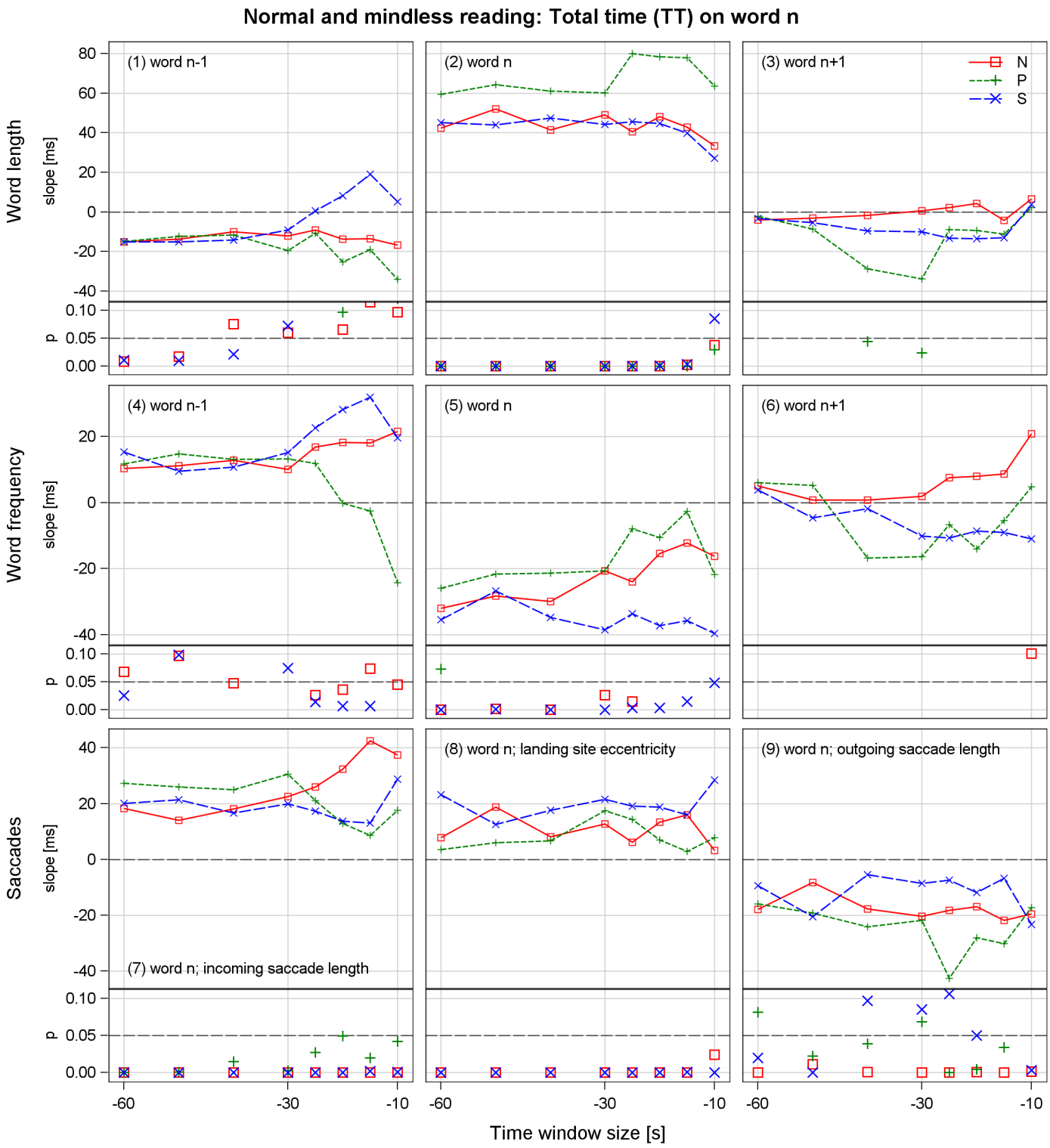


Figure 4.26: Effects (slopes) of word length (in character spaces), word frequency (natural log; SUBTLEX, Brysbaert & New, 2009), length of an incoming and outgoing saccade (in character spaces), and landing site eccentricity (in character spaces) of the previous (n-1), current (n), and next (n+1) word on total time (*em.oa.tt*) on the current word for normal reading (N), probe-caught mindless reading (P), and self-caught mindless reading (S) and for time windows of 10–60s (model (4.10)). Data shown in Table 4.16.

Table 4.14: Effects (slopes; p-values in brackets) of word length (in character spaces), word frequency (natural log; SUBTLEX, Brysbaert & New, 2009), length of an incoming and outgoing saccade (in character spaces), and landing site eccentricity (in character spaces) of the previous (n-1), current (n), and next (n+1) word on gaze duration ($em.fp.gd$) on the current word for normal reading (N), probe-caught mindless reading (P), and self-caught mindless reading (S) and for time windows of 4–60s (model (4.10)). A p-value of .000 means $.0001 < p < .001$, while 0 denotes $p < .0001$. Data plotted in Figure 4.24.

r	t	Word length ($w.len$)						Word frequency ($w.frq$)						Saccade ($sacc$)					
		n-1		n		n+1		n-1		n		n+1		$in.len$		$land.ecc$		$out.len$	
N	4	19	(.300)	46	^{b}(.008)	-9	(.451)	29	^{a}(.028)	5	(.612)	-11	(.411)	12	^{a}(.013)	18	^{a}(.011)	-21	(.134)
	6	20	(.156)	24	(.185)	-17	(.181)	33	^{c}(.001)	-9	(.655)	-6	(.582)	21	^{d}(0)	20	^{d}(0)	-15	^{a}(.037)
	8	11	(.572)	16	(.449)	-1	(.837)	27	^{b}(.002)	-19	(.155)	8	(.486)	20	^{d}(0)	6	^{c}(.000)	-10	(.064)
	10	1	(.770)	24	(.066)	-1	(.900)	19	^{b}(.008)	-15	(.202)	0	(.917)	20	^{d}(0)	6	^{c}(.000)	-11	^{a}(.018)
	15	-1	(.455)	21	^{a}(.050)	0	(.915)	14	^{a}(.028)	-21	^{a}(.013)	-1	(.689)	20	^{d}(0)	9	^{d}(0)	-12	^{b}(.003)
	20	-1	(.432)	20	^{a}(.019)	2	(.641)	9	(.065)	-22	^{b}(.001)	0	(.978)	15	^{d}(0)	11	^{d}(0)	-10	^{b}(.007)
	25	1	(.650)	16	(.061)	2	(.526)	8	(.065)	-25	^{c}(.000)	-1	(.680)	13	^{d}(0)	7	^{d}(0)	-11	^{c}(.001)
	30	-1	(.458)	19	^{a}(.010)	-1	(.999)	3	(.342)	-21	^{c}(.000)	-3	(.507)	14	^{d}(0)	9	^{d}(0)	-11	^{c}(.000)
	40	1	(.617)	20	^{b}(.002)	1	(.680)	6	(.086)	-21	^{c}(.000)	0	(.873)	11	^{d}(0)	7	^{d}(0)	-8	^{b}(.008)
	50	-6	(.064)	34	^{d}(0)	-1	(.773)	6	(.092)	-12	(.066)	-4	(.330)	12	^{d}(0)	10	^{d}(0)	-4	(.102)
	60	-4	(.122)	23	^{d}(0)	2	(.505)	6	^{a}(.045)	-18	^{c}(.000)	2	(.884)	13	^{d}(0)	7	^{d}(0)	-8	^{b}(.001)
P	4	-28	(.157)	18	(.550)	5	(.626)	26	(.524)	-16	(.468)	15	(.498)	73	^{c}(.000)	1	(.447)	-20	(.058)
	6	-34	(.062)	24	(.478)	-7	(.919)	-14	(.562)	6	(.816)	18	(.407)	53	^{c}(.000)	6	(.223)	-8	(.147)
	8	-24	(.113)	24	(.371)	10	(.615)	-21	(.460)	-13	(.661)	15	(.399)	29	^{b}(.001)	10	(.244)	-23	(.068)
	10	-20	(.200)	20	(.438)	-7	(.609)	-14	(.619)	-17	(.389)	-7	(.729)	23	^{b}(.005)	10	(.315)	-21	(.133)
	15	-6	(.444)	50	^{b}(.002)	-3	(.768)	12	(.260)	11	(.390)	-2	(.687)	16	^{c}(.001)	2	(.186)	-25	^{a}(.017)
	20	-16	(.129)	44	^{b}(.002)	-5	(.654)	3	(.574)	-1	(.935)	-4	(.569)	14	^{b}(.004)	1	(.346)	-19	^{a}(.010)
	25	-4	(.485)	45	^{c}(.000)	-9	(.436)	5	(.408)	-2	(.962)	-4	(.545)	18	^{c}(.001)	1	(.155)	-24	^{b}(.003)
	30	-7	(.256)	33	^{a}(.011)	-15	(.150)	8	(.274)	-10	(.438)	-1	(.851)	17	^{c}(.000)	1	(.091)	-12	(.190)
	40	-3	(.513)	29	^{a}(.015)	-18	^{a}(.038)	9	(.220)	-10	(.355)	-5	(.708)	23	^{d}(0)	-3	(.719)	-10	(.464)
	50	-1	(.604)	33	^{b}(.003)	-5	(.548)	9	(.185)	-11	(.278)	5	(.540)	22	^{d}(0)	-2	(.189)	-9	(.190)
	60	-2	(.420)	31	^{b}(.004)	2	(.791)	9	(.226)	-12	(.212)	7	(.358)	25	^{d}(0)	1	^{a}(.046)	-7	(.443)
S	4	-7	(.618)	29	(.141)	14	(.311)	-12	(.773)	-1	(.982)	12	(.434)	10	(.202)	-1	(.393)	-34	^{a}(.015)
	6	2	(.941)	20	(.178)	-3	(.743)	9	(.314)	-1	(.919)	5	(.521)	15	(.053)	9	(.152)	-10	(.412)
	8	2	(.760)	11	(.535)	11	(.201)	18	^{a}(.041)	-23	(.151)	3	(.624)	24	^{d}(0)	12	^{b}(.001)	-7	(.478)
	10	10	(.653)	18	(.217)	5	(.342)	19	^{a}(.027)	-25	(.084)	-4	(.824)	25	^{d}(0)	13	^{c}(.000)	-6	(.099)
	15	16	(.183)	27	^{b}(.008)	-3	(.865)	22	^{b}(.003)	-16	(.132)	-7	(.401)	13	^{d}(0)	9	^{c}(.000)	-5	(.056)
	20	5	(.983)	37	^{c}(.000)	-1	(.655)	19	^{b}(.002)	-12	(.218)	-4	(.599)	16	^{d}(0)	14	^{d}(0)	-5	^{a}(.028)
	25	1	(.532)	36	^{d}(0)	-2	(.672)	12	^{a}(.022)	-12	(.129)	-3	(.596)	13	^{d}(0)	14	^{d}(0)	-2	(.123)
	30	0	(.458)	34	^{d}(0)	-1	(.616)	11	^{a}(.020)	-19	^{a}(.011)	-7	(.215)	14	^{d}(0)	13	^{d}(0)	-5	(.124)
	40	-5	(.109)	33	^{d}(0)	1	(.608)	9	^{a}(.040)	-16	^{a}(.015)	-2	(.683)	12	^{d}(0)	10	^{d}(0)	-5	(.171)
	50	-1	(.313)	22	^{c}(.000)	1	(.629)	7	(.058)	-19	^{c}(.000)	-1	(.742)	13	^{d}(0)	7	^{d}(0)	-7	^{b}(.001)
	60	-4	(.098)	32	^{d}(0)	-1	(.859)	9	^{a}(.024)	-15	^{b}(.009)	-4	(.401)	13	^{d}(0)	12	^{d}(0)	-5	(.161)

Significance level indication: $a - 0.05$, $b - 0.01$, $c - 0.001$, $d - 0.0001$

Table 4.15: Effects (slopes; p-values in brackets) of word length (in character spaces), word frequency (natural log; SUBTLEX, Brysbaert & New, 2009), length of an incoming and outgoing saccade (in character spaces), and landing site eccentricity (in character spaces) of the previous (n-1), current (n), and next (n+1) word on single fixation duration ($em.fp.sfd$) on the current word for normal reading (N), probe-caught mindless reading (P), and self-caught mindless reading (S) and for time windows of 4–60s (model (4.10)). A p-value of .000 means $.0001 < p < .001$, while 0 denotes $p < .0001$. Data plotted in Figure 4.25.

r	t	Word length ($w.len$)						Word frequency ($w.frq$)						Saccade ($sacc$)					
		n-1		n		n+1		n-1		n		n+1		$in.len$		$land.ecc$		$out.len$	
N	4	23	(.108)	20	(.186)	-13	(.218)	18	(.105)	9	(.456)	-13	(.276)	17	^{a}(.015)	2	(.840)	-14	(.284)
	6	15	(.117)	3	(.961)	-11	(.165)	17	^{a}(.035)	-1	(.919)	-7	(.492)	24	^{c}(.000)	3	(.727)	-11	(.356)
	8	7	(.491)	-3	(.509)	-1	(.859)	9	(.212)	-4	(.776)	4	(.566)	19	^{c}(.000)	-7	(.557)	-8	(.465)
	10	2	(.985)	1	(.775)	0	(0000)	5	(.416)	-3	(.754)	-1	(.999)	19	^{d}(0)	-6	(.665)	-8	(.405)
	15	1	(.915)	5	(.675)	-3	(.606)	6	(.251)	-2	(.775)	-2	(.669)	18	^{d}(0)	-6	(.734)	-6	(.051)
	20	4	(.568)	1	(.939)	2	(.853)	4	(.342)	-9	(.102)	-2	(.764)	14	^{c}(.000)	-2	(.548)	-9	(.057)
	25	4	(.483)	-2	(.481)	2	(.688)	5	(.250)	-11	^{a}(.023)	-2	(.632)	11	^{c}(.000)	-5	(.455)	-10	^{b}(.004)
	30	4	(.524)	-4	(.225)	0	(.949)	-1	(.961)	-11	^{a}(.015)	1	(.928)	13	^{d}(0)	-5	(.483)	-9	^{b}(.003)
	40	3	(.568)	-4	(.174)	0	(.992)	0	(.887)	-13	^{b}(.002)	1	(.724)	12	^{d}(0)	-7	(.135)	-5	(.149)
	50	0	(.774)	8	(.077)	-3	(.411)	2	(.509)	-2	(.905)	-3	(.446)	11	^{d}(0)	-9	^{b}(.004)	2	(.357)
	60	1	(.971)	0	(.659)	-1	(.692)	-1	(.948)	-9	^{a}(.020)	0	(.889)	12	^{d}(0)	-6	(.138)	-5	^{a}(.040)
P	4	-6	(.670)	-3	(.874)	10	(.479)	56	(.063)	13	(.616)	13	(.599)	72	^{b}(.001)	7	(.572)	-13	(.203)
	6	4	(.942)	-13	(.477)	-7	(.715)	46	^{a}(.027)	19	(.393)	12	(.467)	65	^{c}(.000)	10	(.593)	-8	(.380)
	8	-10	(.509)	-1	(.922)	14	(.431)	8	(.540)	3	(.832)	15	(.260)	42	^{b}(.002)	4	(.725)	-12	(.337)
	10	-8	(.634)	-3	(.837)	7	(.672)	11	(.377)	-3	(.847)	-2	(.942)	24	^{a}(.018)	6	(.619)	-10	(.400)
	15	-7	(.447)	7	(.685)	5	(.801)	10	(.322)	4	(.707)	1	(.939)	16	^{b}(.009)	-3	(.952)	-12	(.303)
	20	-10	(.253)	5	(.676)	1	(.976)	9	(.263)	-4	(.771)	-4	(.574)	16	^{a}(.018)	1	(.989)	-10	(.084)
	25	0	(.845)	15	(.100)	-4	(.547)	8	(.261)	7	(.339)	-6	(.378)	15	^{a}(.011)	-1	(.845)	-17	^{b}(.009)
	30	-3	(.600)	12	(.218)	-6	(.409)	8	(.225)	6	(.390)	-4	(.623)	15	^{b}(.003)	-1	(.912)	-11	(.065)
	40	0	(.834)	8	(.408)	-13	(.069)	8	(.259)	0	(.926)	-8	(.343)	18	^{b}(.001)	-4	(.546)	-11	(.084)
	50	4	(.585)	10	(.248)	-3	(.584)	10	(.160)	2	(.707)	1	(.775)	21	^{c}(.000)	-7	(.176)	-10	(.102)
	60	0	(.916)	6	(.478)	-2	(.679)	4	(.565)	3	(.638)	0	(.843)	23	^{c}(.000)	-4	(.217)	-7	(.428)
S	4	-10	(.538)	-1	(.719)	-1	(.693)	-15	(.333)	2	(.916)	13	(.423)	12	(.576)	-20	(.065)	-7	(.943)
	6	-1	(.870)	1	(.868)	-13	(.325)	-6	(.545)	10	(.369)	2	(.493)	11	(.355)	-19	^{b}(.003)	8	(.069)
	8	1	(.967)	-5	(.418)	0	(.832)	5	(.355)	-3	(.908)	6	(.451)	25	^{b}(.004)	-12	(.131)	5	(.125)
	10	15	(.122)	-4	(.425)	-3	(.682)	6	(.265)	-8	(.465)	-5	(.742)	13	^{a}(.014)	-16	(.089)	3	(.242)
	15	16	^{a}(.029)	10	(.283)	-2	(.697)	9	(.097)	1	(.774)	-5	(.569)	9	^{a}(.021)	-12	(.069)	4	(.200)
	20	8	(.185)	13	(.079)	-3	(.508)	11	^{a}(.034)	-1	(.929)	-5	(.522)	12	^{b}(.001)	-6	(.136)	1	(.194)
	25	4	(.520)	9	(.196)	-3	(.458)	5	(.205)	-4	(.651)	-6	(.379)	10	^{c}(.001)	-7	(.178)	1	(.280)
	30	4	(.416)	10	(.112)	-1	(.682)	5	(.166)	-6	(.298)	-6	(.278)	11	^{c}(.001)	-6	(.098)	1	(.339)
	40	2	(.685)	8	(.163)	-2	(.464)	5	(.143)	-5	(.350)	-3	(.600)	10	^{c}(.000)	-8	^{b}(.004)	-1	(.407)
	50	3	(.574)	-2	(.382)	0	(.996)	1	(.718)	-11	^{b}(.007)	0	(.960)	12	^{d}(0)	-5	(.374)	-5	^{a}(.037)
	60	1	(.979)	9	(.052)	-3	(.281)	5	(.153)	-2	(.831)	-3	(.609)	12	^{d}(0)	-7	^{b}(.002)	1	(.193)

Significance level indication: $a - 0.05$, $b - 0.01$, $c - 0.001$, $d - 0.0001$

Table 4.16: Effects (slopes; p-values in brackets) of word length (in character spaces), word frequency (natural log; SUBTLEX, Brysbaert & New, 2009), length of an incoming and outgoing saccade (in character spaces), and landing site eccentricity (in character spaces) of the previous (n-1), current (n), and next (n+1) word on total time ($em.oa.tt$) on the current word for normal reading (N), probe-caught mindless reading (P), and self-caught mindless reading (S) and for time windows of 4–60s (model (4.10)). A p-value of .000 means $.0001 < p < .001$, while 0 denotes $p < .0001$. Data plotted in Figure 4.26.

r	t	Word length ($w.len$)						Word frequency ($w.frq$)						Saccade ($sacc$)					
		n-1		n		n+1		n-1		n		n+1		$in.len$		$land.ecc$		$out.len$	
N	4	6	(.860)	42	(.050)	1	(.947)	32	^{a}(.044)	11	(.438)	8	(.673)	21	^{a}(.012)	21	^{b}(.009)	-29	^{a}(.036)
	6	17	(.379)	20	(.450)	-10	(.740)	42	^{c}(.001)	-10	(.777)	3	(.979)	32	^{d}(0)	23	^{d}(0)	-21	^{c}(.001)
	8	0	(.692)	19	(.419)	11	(.252)	36	^{b}(.004)	-20	(.256)	22	(.145)	40	^{d}(0)	13	^{b}(.001)	-16	^{b}(.004)
	10	-17	(.096)	33	^{a}(.038)	7	(.381)	22	^{a}(.045)	-16	(.303)	21	(.101)	37	^{d}(0)	3	^{a}(.024)	-19	^{b}(.001)
	15	-13	(.115)	43	^{b}(.003)	-4	(.769)	18	(.074)	-12	(.438)	9	(.581)	43	^{d}(0)	16	^{d}(0)	-22	^{c}(.000)
	20	-14	(.066)	48	^{c}(.000)	4	(.489)	18	^{a}(.036)	-15	(.187)	8	(.479)	32	^{d}(0)	13	^{d}(0)	-17	^{c}(.001)
	25	-9	(.147)	40	^{c}(.000)	2	(.596)	17	^{a}(.026)	-24	^{a}(.015)	8	(.542)	26	^{d}(0)	6	^{c}(.000)	-18	^{c}(.000)
	30	-12	(.060)	49	^{d}(0)	1	(.761)	10	(.129)	-21	^{a}(.026)	2	(.998)	23	^{d}(0)	13	^{d}(0)	-20	^{d}(0)
	40	-10	(.076)	42	^{d}(0)	-2	(.919)	13	^{a}(.048)	-30	^{c}(.000)	1	(.978)	18	^{d}(0)	8	^{d}(0)	-18	^{c}(.000)
	50	-14	^{a}(.017)	52	^{d}(0)	-3	(.858)	11	(.097)	-28	^{b}(.002)	1	(.998)	14	^{d}(0)	19	^{d}(0)	-8	^{a}(.011)
	60	-15	^{b}(.008)	42	^{d}(0)	-4	(.804)	10	(.068)	-32	^{d}(0)	5	(.620)	18	^{d}(0)	8	^{d}(0)	-18	^{d}(0)
P	4	-20	(.375)	15	(.538)	-5	(.903)	47	(.188)	-20	(.507)	11	(.755)	83	^{c}(.000)	36	(.076)	-47	^{b}(.006)
	6	-34	(.117)	47	(.183)	1	(.811)	-9	(.852)	3	(.885)	24	(.351)	41	^{b}(.001)	38	^{a}(.049)	-48	^{b}(.007)
	8	-41	(.051)	32	(.293)	24	(.392)	-25	(.492)	-26	(.423)	43	(.064)	33	^{b}(.002)	32	(.152)	-38	(.058)
	10	-34	(.122)	64	^{a}(.029)	2	(.919)	-24	(.410)	-22	(.421)	5	(.794)	18	^{a}(.042)	8	(.961)	-17	(.287)
	15	-19	(.187)	78	^{c}(.000)	-11	(.462)	-3	(.908)	-3	(.958)	-5	(.645)	9	^{a}(.020)	3	(.611)	-30	^{a}(.034)
	20	-25	(.097)	78	^{d}(0)	-9	(.528)	0	(.760)	-11	(.673)	-14	(.311)	13	^{a}(.050)	7	(.794)	-28	^{b}(.004)
	25	-11	(.381)	80	^{d}(0)	-9	(.529)	12	(.317)	-8	(.763)	-7	(.529)	21	^{a}(.027)	14	(.473)	-43	^{c}(.000)
	30	-19	(.130)	60	^{c}(.000)	-34	^{a}(.024)	13	(.315)	-21	(.208)	-16	(.259)	31	^{b}(.002)	17	(.282)	-22	(.069)
	40	-12	(.341)	61	^{c}(.000)	-29	^{a}(.045)	13	(.298)	-21	(.164)	-17	(.294)	25	^{a}(.015)	7	(.925)	-24	^{a}(.039)
	50	-12	(.253)	64	^{d}(0)	-9	(.537)	15	(.220)	-22	(.163)	5	(.759)	26	^{c}(.001)	6	(.443)	-19	^{a}(.022)
	60	-15	(.126)	59	^{c}(.000)	-2	(.857)	12	(.292)	-26	(.073)	6	(.666)	27	^{d}(0)	4	(.225)	-16	(.082)
S	4	-9	(.483)	40	^{a}(.038)	1	(.612)	-20	(.338)	-4	(.852)	2	(.885)	4	(.603)	19	(.056)	-38	^{b}(.005)
	6	1	(.793)	28	(.135)	-17	(.431)	4	(.791)	-4	(.922)	-7	(.595)	18	(.120)	19	(.079)	-13	(.280)
	8	1	(.558)	18	(.242)	4	(.547)	20	(.185)	-36	(.117)	-6	(.833)	29	^{c}(.001)	31	^{c}(.000)	-20	^{a}(.038)
	10	5	(.786)	27	(.085)	4	(.544)	20	(.158)	-40	^{a}(.048)	-11	(.482)	29	^{c}(.000)	28	^{c}(.000)	-23	^{b}(.003)
	15	19	(.374)	40	^{b}(.003)	-13	(.455)	32	^{b}(.007)	-36	^{a}(.015)	-9	(.453)	13	^{b}(.002)	16	^{c}(.001)	-7	(.140)
	20	8	(.979)	45	^{b}(.001)	-14	(.363)	28	^{b}(.006)	-37	^{b}(.004)	-9	(.447)	14	^{d}(0)	19	^{d}(0)	-12	^{a}(.050)
	25	1	(.528)	46	^{c}(.000)	-13	(.363)	23	^{a}(.014)	-34	^{b}(.003)	-11	(.259)	17	^{d}(0)	19	^{d}(0)	-7	(.106)
	30	-9	(.072)	44	^{c}(.000)	-10	(.543)	15	(.075)	-38	^{c}(.000)	-10	(.249)	20	^{d}(0)	22	^{d}(0)	-8	(.085)
	40	-14	^{a}(.021)	47	^{d}(0)	-10	(.569)	11	(.152)	-35	^{c}(.001)	-2	(.752)	17	^{d}(0)	18	^{d}(0)	-5	(.097)
	50	-15	^{a}(.010)	44	^{d}(0)	-5	(.662)	10	(.098)	-27	^{c}(.001)	-5	(.425)	21	^{d}(0)	13	^{d}(0)	-20	^{d}(0)
	60	-15	^{a}(.011)	45	^{d}(0)	-3	(.919)	15	^{a}(.026)	-35	^{d}(0)	4	(.649)	20	^{d}(0)	23	^{d}(0)	-9	^{a}(.020)

Significance level indication: $a - 0.05$, $b - 0.01$, $c - 0.001$, $d - 0.0001$

Conclusions and Discussion

In my analysis of the overall trend I investigated if lag and successor effects are present in the eye-movement record from my natural reading task. The two most anticipated effect I have found are that of the strong positive immediate effect of a word's length and strong immediate negative effect of a word's frequency. Reports of similar effects are prevalent in the literature (e.g., see Rayner, 1998, 2009). The only lag and successor effects I have found were small and restricted to word length. Furthermore, I have not found evidence for the most contentious effect (cf., Kliegl, 2007; Rayner et al., 2007), the frequency successor effect.

I have also found that the current word's processing times were inflated proportionally to the length of incoming saccades. This is most likely due to the limited parafoveal preview associated with longer saccades (Radach & Heller, 2000; Vitu et al., 2001). Furthermore, faster processing times associated with landing sites being to the left of center of words are evidence of preferred viewing position (McConkie et al., 1988). The only problematic relationship I have come across is that of longer inspection times being followed by shorter outgoing saccades. This effect is counter-intuitive in that if longer processing times grant more parfoveal preview, then they should be followed by longer saccades. However, processing difficult words has been found to restrict the amount of parafoveal preview (Henderson & Ferreira, 1990; Kennison & Clifton, 1995) and therefore if the elevated processing times I have observed were due to word difficulty then shorter outgoing saccades would in fact be expected.

In my investigation of lag and successor effects during normal and mindless reading, I have found evidence for shallower text processing during bouts of mind-wandering. More specifically, I have found that the immediacy word-length effect was stronger and the immediacy word-frequency effect was weaker (or even absent) during mindless reading. These results indicate that during moments of attention decoupling, readers rely (potentially exclusively) on perceptual instead of lexical (and, by natural extension, linguistic) properties of words being read. This eventuality has been suggested earlier by (Reichle, Reineberg, & Schooler, 2010).

Similarly to the overall trend analysis, I have found weak evidence for the lag and successor word-length effects during both normal and mindless reading. Unlike that analysis though, I have also observed the "anti-spillover" effect (i.e., positive word-frequency lag effect). I call it this way because spillover effects (Rayner & Duffy, 1986) reported in the literature go in the opposite direction from the one I have found in my data and are a result of some portion of processing of the previous word that completes only after the next word has already been fixated. While I cannot provide a convincing explanation of this finding, it could be a result of the way I sampled eye movement data, selected word triplet, or both of these.

Finally, the comparative normal-vs-mindless reading analysis informs my modeling attempts. That is, due to the general absence of strong lag and successor effects I am inclined to focusing on only the current word. Incidentally, by accounting for the immediacy effects only I can use more data by completely sidestepping the issues of triplets selection and the associated inevitable data loss (Rayner et al., 2003).

4.11 Word Effects

Motivation

In this section, as a final step before moving on to detecting mindless reading, I further scrutinize the way it can manifest itself by analyzing its influence on a number of eye-movement variables which have been investigated by reading researchers. My primary goal is to discover which of these variables are likely to be affected by the changing attentional state of a reader and consequently using them in my modeling work. In the previous section, I have weighed the evidence for and against the lag and successor effects in both normal and mindless reading. Because I have found these effects to be weak at best, below, I focus on the current word only. Like in the previous section, I use two word-level variables: Length ($w.len$) and frequency ($w.frq$).

Table 4.11 lists all variables I focus on in this section (along with certain details pertinent to the statistical models I fit, as described later). As is clear from that list, I focus on first-pass variables. The reason for this is that second-pass variables are less useful for detecting mindless reading because ideally the reader will be caught mind-wandering already during first-pass reading and that is what I aim for in this work. Moreover, from a purely pragmatic point of view, because second-pass reading can include many instances of re-reading it would difficult to clearly define which eye-movements were to be used. That is, would it be the first second-pass reading, the second second-pass reading, or perhaps the tenth second-pass reading? I choose not to deal with this sort of ambiguities. There is only one exception to this rule (and every good rule should have one). Namely, because task-evoked pupillary response does not reliably occur for any one instance of cognitive effort (Beatty, 1982; Beatty & Lucero-Wagoner, 2000), I compute pupil diameter ($em.oa.pd$) based on fixations from both first- and second-pass reading.

Models and Data

To investigate if values of eye-movement variables were different for words with different length and frequency during normal and mindless reading I fitted a series of models, each given as

$$
\begin{aligned}
y_{ijk}|s_i &\sim Q,\\
\mu_{ijk} &= g(\eta_{ijk}),\\
\eta_{ijk} &= \beta_0 + \beta_1(w.len_{ij}) + \beta_2(w.frq_{ij}) + \beta_3(r_k)\\
&\quad + \beta_4(w.len_{ij} \times w.frq_{ij}) + \beta_5(w.len_{ij} \times r_k)\\
&\quad + \beta_6(w.frq_{ij} \times r_k) + \beta_7(w.len_{ij} \times w.frq_{ij} \times r_k)\\
&\quad + s_i,\\
s_i &\sim iid\ \text{Normal}(0, \sigma_s^2),\\
i &\in \{1, 2, \ldots, 99\},\\
k &\in \{1, 2, 3\},
\end{aligned}
\tag{4.11}
$$

where y is the response variable, μ is the mean of the distribution Q, s is the subject random effect, i is the subject index, j is the reading mode (i.e., N, P, and S) index, and k is the word index. η is the linear predictor component of the GLMM. As a reminder, the link function $g(\cdot)$ is used to associate the mean of the response variable distribution Q with the linear predictor. The inverse link function $g^{-1}(\cdot)$ is used to perform the opposite transformation. The selection of the link function is therefore intimately related to the distribution of the response variable. The distributions and the corresponding link functions I use are given in Table 4.17. Note, that I do not provide forms of the variance function; instead, I refer the interested reader to the manual of the SAS GLIMMIX procedure for information on variance functions for each response distribution I use. All p-values I report in this section have been adjusted by simulation. Just like in the previous section, in this analysis I used time windows of 4–60s as sampling frames.

Results

Figures 4.27–4.39 show the results of fitting model 4.11 to each of the response variables listed in Table 4.17. They share the same layout template which is divided into 12 subplots. The top portion of each of these subplots shows means of the response variables and the bottom portion shows p-values associated with tests of differences between these means. These tests of differences check if the respective difference is significantly far from zero and they compare either (a) short and long words (subplots 1 and 4), (b) frequent and infrequent words (subplot 2 and 4), or (d) normal reading, probe-caught mindless reading, and self-caught mindless reading (subplots 3 and 5–12). The subplots are numbered and show the following data point series:

1. Means for short ($w.len = 0$) and long words ($w.len = 1$)
2. Means for infrequent ($w.frq = 0$) and frequent words ($w.frq = 1$)
3. Means for normal reading (r = N), probe-caught mindless reading (r = P), and self-caught mindless reading (r = S)
4. Means for short infrequent (00), short frequent (01), long infrequent (10), and long frequent (11) words ($w.len \times w.frq$)
5. Means for normal and mindless reading for short words
6. Means for normal and mindless reading for long words
7. Means for normal and mindless reading for infrequent words

Table 4.17: Eye-movement variables I investigate in this section along with the respective distributions (Q) and link functions ($h(\cdot)$) I use when fitting model (4.11).

Variable		Model	
Long name	Short name	Q	$g(\cdot)$
Inspection durations [ms]			
Gaze duration	$em.gd$		
Single fixation duration	$em.sfd$	Lognormal	Log
First-fixation duration	$em.ffd$		
Inspection probabilities			
Prob. of skipping	$em.p0$		
Prob. of one fixation	$em.p1$	Binomial	Logit
Prob. of two or more fixations	$em.p2$		
Other probabilities			
Prob. of regression	$em.pr$	Binomial	Logit
Prob. of landing site left off-center	$em.plsloc$		
Saccade amplitudes [char]			
Forward incoming saccade amp. ($\leq$ 15 char)[1)]	$em.fisa$		
Forward outgoing saccade amp. ($\leq$ 15 char)[1)]	$em.fosa$	Normal	Id
Regressive outgoing saccade amp. ($\leq$ 30 char)[2)]	$em.rosa$		
Saccade other			
Landing site eccentricity ($\leq$ 15 char)[3)]	$em.lse$	Normal	Id
Pupilometry [%]			
Pupil diameter	$em.pd$	Normal	Id

[1)] Saccade longer than 15 characters would bring into the fovea text that fell outside the perceptual span.

[2)] I consider only regressions no longer than 30 characters.

8. Means for normal and mindless reading for frequent words
9. Means for normal and mindless reading for short infrequent words
10. Means for normal and mindless reading for short frequent words
11. Means for normal and mindless reading for long infrequent words
12. Means for normal and mindless reading for long frequent words

I start my analysis with inspection duration measures. Figure 4.27 shows the results for gaze duration ($em.fp.gd$). As expected, gaze duration is inflated for long words compared to short words, and inflated for infrequent words compared to frequent words, with short frequent words being processed the fastest and long infrequent words being processed the slowest. Furthermore, the patterns of p-values suggest that gaze durations are shorter during normal reading as compared to both probe-caught and self-caught mindless reading in time windows wider than 10s. This gaze-duration gap is the widest for long infrequent words. No evidence for the difference between the two kinds of mindless reading is present.

Figure 4.28 shows the results for single fixation duration ($em.fp.sfd$). Unlike gaze duration, single-fixations duration does not vary with word length, but does vary with word frequency and in an anticipated way with infrequent words being associated with longer single-fixation durations. The patterns of p-values suggests that single-fixation durations are shorter during normal reading as compared to both kinds of mindless reading. Interestingly, while in the case of gaze duration this regularity was present for long infrequent words only, single-fixation durations vary with the attentional state of a reader in a more complex way involving long frequent words as well. The story for first fixation duration ($em.fp.ffd$; Figure 4.29) is essentially the same, with the exception that the evidence for separation between normal reading and probe-caught

mindless reading is weaker than for single fixation duration.

I now move to discussing inspection probabilities. Figure 4.30 shows the results for probability of skipping. As expected, short frequent words are noticeably more likely to be skipped than long infrequent words. Additionally, there is a reasonable amount of evidence that more words are skipped during episodes of self-caught mindless reading compared to normal reading within about 20 seconds prior to a reader catching themselves mind-wandering. This difference is restricted to long frequent words and the largest separation is visible for the 10s time window. This result is quite interesting as it implies that relatively close to becoming aware of one's mind-wandering a reader starts skipping words which are otherwise quite likely to end up in a saccadic crosshair.

Figure 4.31 shows the results for probability of one fixation. As expected, word length and frequency had the opposite effects than they had on skipping behavior – long and infrequent words attracted more single fixations. However, apart from that the remaining evidence is rather scant and indicates that long infrequent words are less likely to receive exactly one fixation during self-caught mindless reading than they are during normal reading. This effect is significant only in the 10–second time window and is related to the symmetrical (and stronger) effect for word skipping I have mentioned earlier.

Figure 4.32 shows the results for probability of two or more fixations ($em.fp.p2$). As anticipated, long infrequent words were the most likely to receive two or more fixations, while short frequent words were the least likely to receive such juicy gifts from a reader. Other than that, however, the is no convincing evidence that mindless reading affects two-or-more fixations behavior.

Figure 4.33 shows the results for probability of a regression ($em.fp.pr$). Evidently, probability of a regression is not modulated by either of the word variables in either normal or mindless reading.

I now proceed to assessing evidence in the three saccade amplitude variables. Note that I limited the amplitude of the incoming and outgoing forward saccades to no more than 15 characters because longer saccades bring into the fovea text that falls outside of the range typically encompassed by the perceptual span characteristic to the English language (McConkie & Rayner, 1976; Rayner & Bertera, 1979; Rayner, Well, & Pollatsek, 1980; Den Buurman & Boersma, 1981; Rayner et al., 1982). Figure 4.34 shows the results for forward incoming saccade amplitude ($em.fp.fisa$). The evidence shows that long infrequent words are targeted from the farthest away while short frequent words are at the other end of the saccadic amplitude spectrum. This is to be expected as landing sites are typically around the center of a word and the center of a longer word will, on average, be farther away from the current fixation location than the center of a short word. Apart from that, there is merely a vestige of a difference between normal reading and self-caught mindless reading for long infrequent words in wide time windows. That difference, however, has a sub-character-space magnitude.

Figure 4.35 shows the results for forward outgoing saccade amplitude ($em.fp.fosa$). It appears that long infrequent words are associated with the longest outgoing saccades. The explanation of this pattern is identical to the one I offered earlier for incoming saccades. Furthermore, like was the case for incoming saccades, there is evidence that outgoing saccades are longer when they are launched during normal reading relative to self-caught mindless reading. Like before, however, the difference is very small and measured in fractions of a character space.

Figure 4.36 shows the results for regressive outgoing saccade amplitude ($em.fp.rosa$). Evidently, longer regressions originate from longer words. This is again most likely due to the fact that the eye needs to be moved back from farther away just because the currently fixated word is longer. Other than that, there is a minimal evidence indicating that regressions launched from long frequent words are shorter when made during self-caught mindless reading as compared to normal reading. While present in only the 15s time window, this difference has a sizable effect size of about two character spaces and could possibly be diagnostic of the reader regaining meta-awareness. Note that because the subjects had access to an entire page of text they could make really long multiline regressions. However, because the proportion of such long regressions was small, I chose to limit the maximum amplitude of all regressions to 30 character spaces and thus cut down on that source of extreme values which are known to affect means (which, naturally, are being modeled here). In an attempt at a rudimentary sensitivity analysis, I found that increasing that threshold to 60 character spaces did not yield different results.

Now, I turn my attention to the remaining two saccade variables. Figure 4.37 shows the results for landing site eccentricity ($em.fp.lse$). Word-length effect is evident with landing sites on longer words biased towards the beginning, as expected. Besides that, no differences between normal and mindless reading exists. The related probability of landing site left-of-center ($em.fp.plsloc$; Figure 4.38) tells an identical simple story.

Finally, the last figure (4.39) shows the results for pupil diameter ($em.oa.pd$). These results reveal an interesting narrative. Is appears that pupil diameter does not respond to variations in word length and word frequency on their own. Evidently though, pupil diameter is smaller when a reader processed long infrequent and short frequent words, but only during probe-caught mindless reading. I return to this observation in the discussion.

Conclusions and Discussion

The three inspection durations measures I have examined above seem promising in discriminating between normal and mindless reading. Because single-fixation and first-fixation durations vary between these two reading reading modes for both frequent and infrequent long words, it seems that modeling word frequency is beneficial. This is further underlined by the observation that long frequent, but not long infrequent, words are more likely to be skipped before a reader catches themselves mind-wandering. This word-skipping behavior is interesting in itself because it indicates that during intervals of mind-wandering eye-movement control becomes disrupted to the point where a reader does not fixate some of the most important words in text. It is plausible that this unexpected skipping marks an onset of the transition between mindless and mindful. Collectively, these results indicate that inspection duration measures should be useful in discovering mind-wandering in readers. Additionally, because of the differences in the probability of word skipping, number of first-pass fixations may also prove fruitful.

Saccade variables provide further evidence that normal and mindless reading can be told apart from each other. However, while amplitudes of both the forward incoming and forward outgoing saccades are sensitive to a reader's attentional state, they vary by no more than a quarter of a character space. Consequently, detecting such a small difference will depend on the spacial resolution of the eye-tracker used. Finally, the regressive outgoing saccade amplitude's difference between normal and self-caught mindless reading is quite large (about 2 character spaces) but evidence for it is present in only one time window and although it could be diagnostic of the reader regaining their meta-awareness its utility for mindless reading detection does not strike me as high. Overall, I do not feel inclined to include saccade variables in my classification models.

Finally, because pupil diameter is known to respond to differences in processing demands (with larger pupil diameters resulting from more cognitive load) the current results suggest that text processing is attenuated during periods of mind-wandering which a reader is unaware of. This supports a hypothesis that probe-caught mindless reading is an instance of deeper mind-wandering.

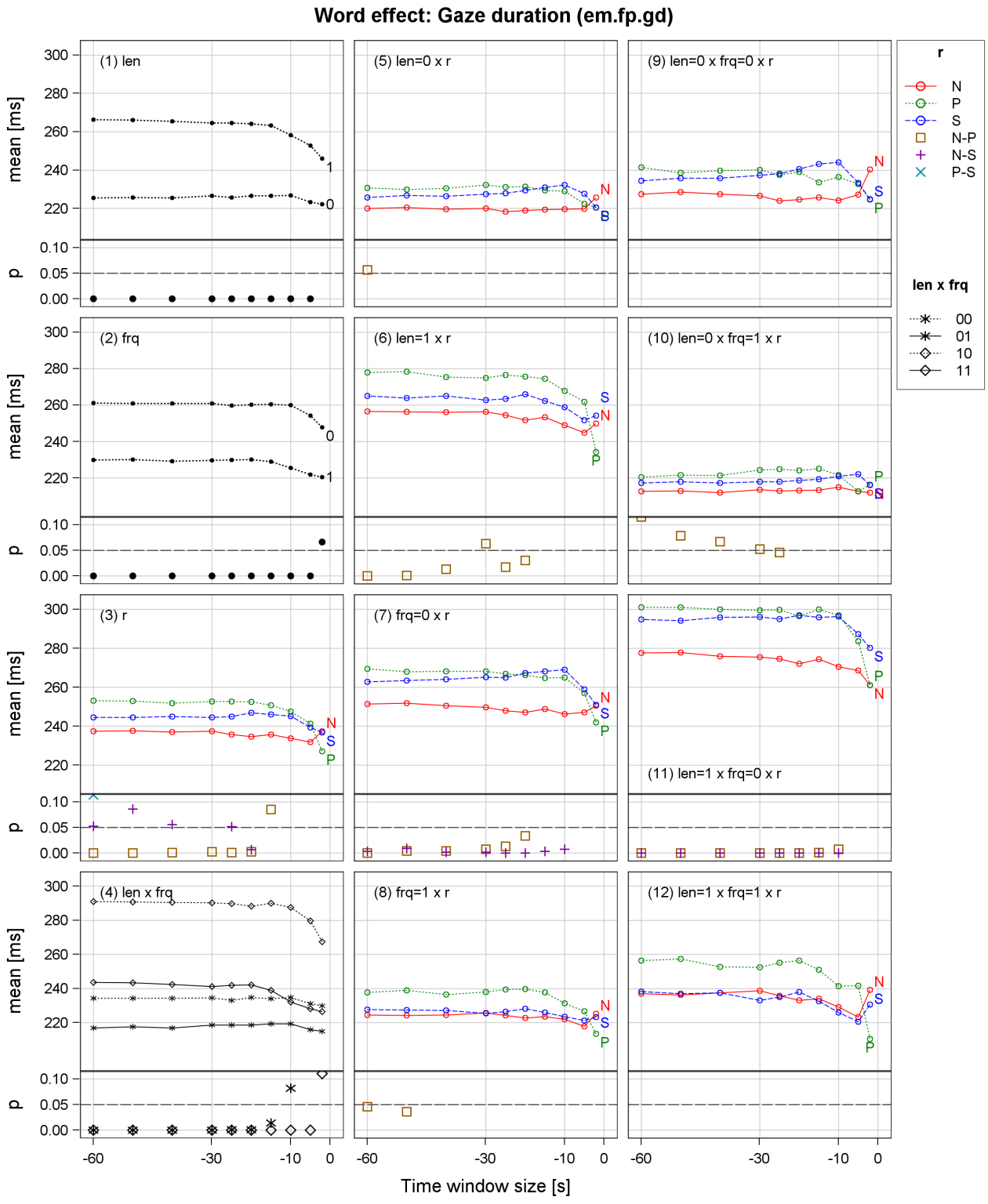


Figure 4.27: Means for main effects and interactions of (1) word length, (2) word frequency, and (3) reading mode (N, P, and S) on gaze duration ($em.fp.gd$). Word length and frequency were dichotomized; length was split on four-or-less characters and frequency on its median. Based on fitting model 4.11 (pg. 63) parameterized as shown in Table 4.17 (pg. 64) to a series of data sets sampled through time windows 4–60s wide. Values plotted are predicted population margins; all p-values were adjusted by simulation.

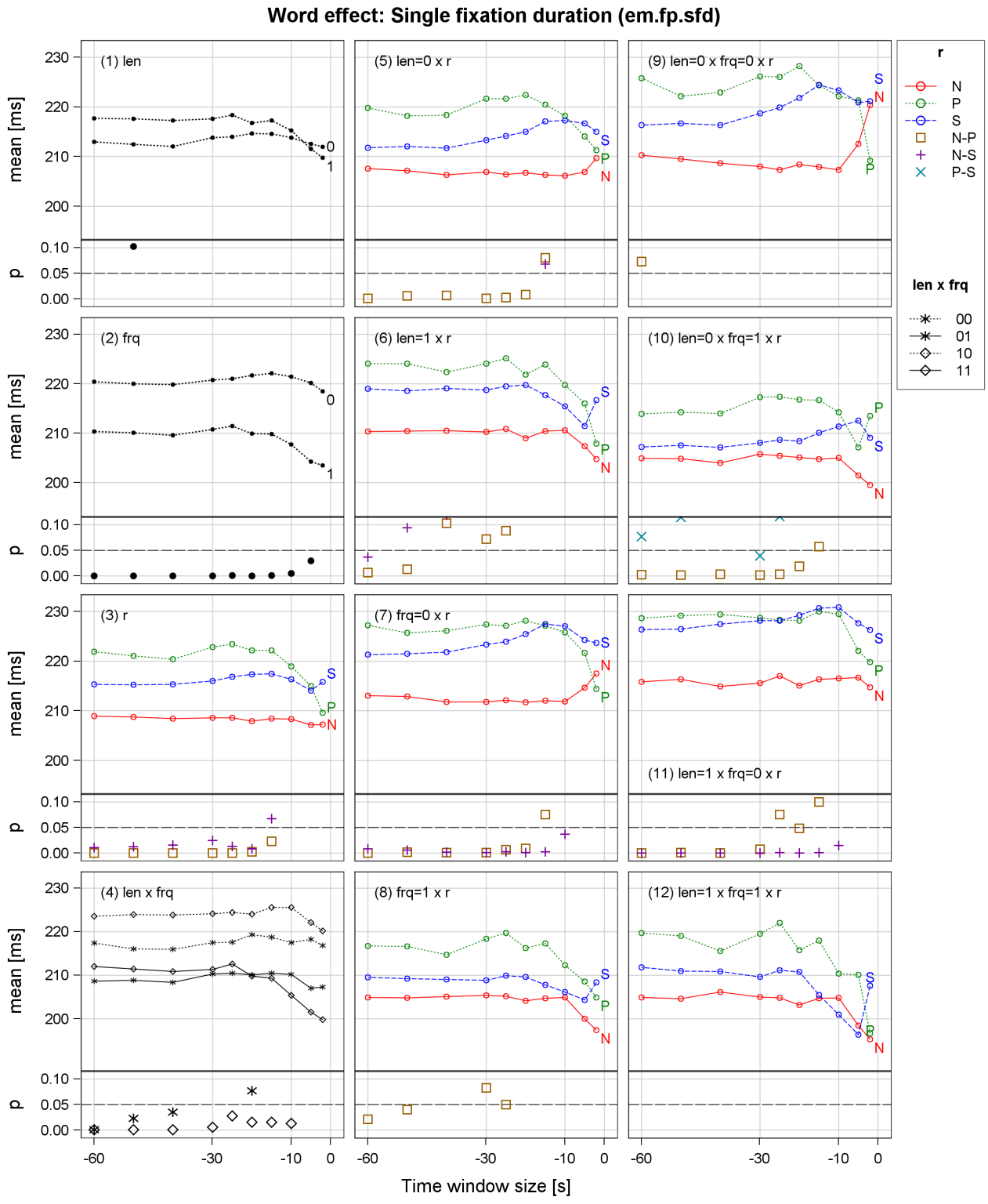


Figure 4.28: Means for main effects and interactions of (1) word length, (2) word frequency, and (3) reading mode (N, P, and S) on single fixation duration ($em.fp.sfd$). Word length and frequency were dichotomized; length was split on four-or-less characters and frequency on its median. Based on fitting model 4.11 (pg. 63) parameterized as shown in Table 4.17 (pg. 64) to a series of data sets sampled through time windows 4–60s wide. Values plotted are predicted population margins; all p-values were adjusted by simulation.

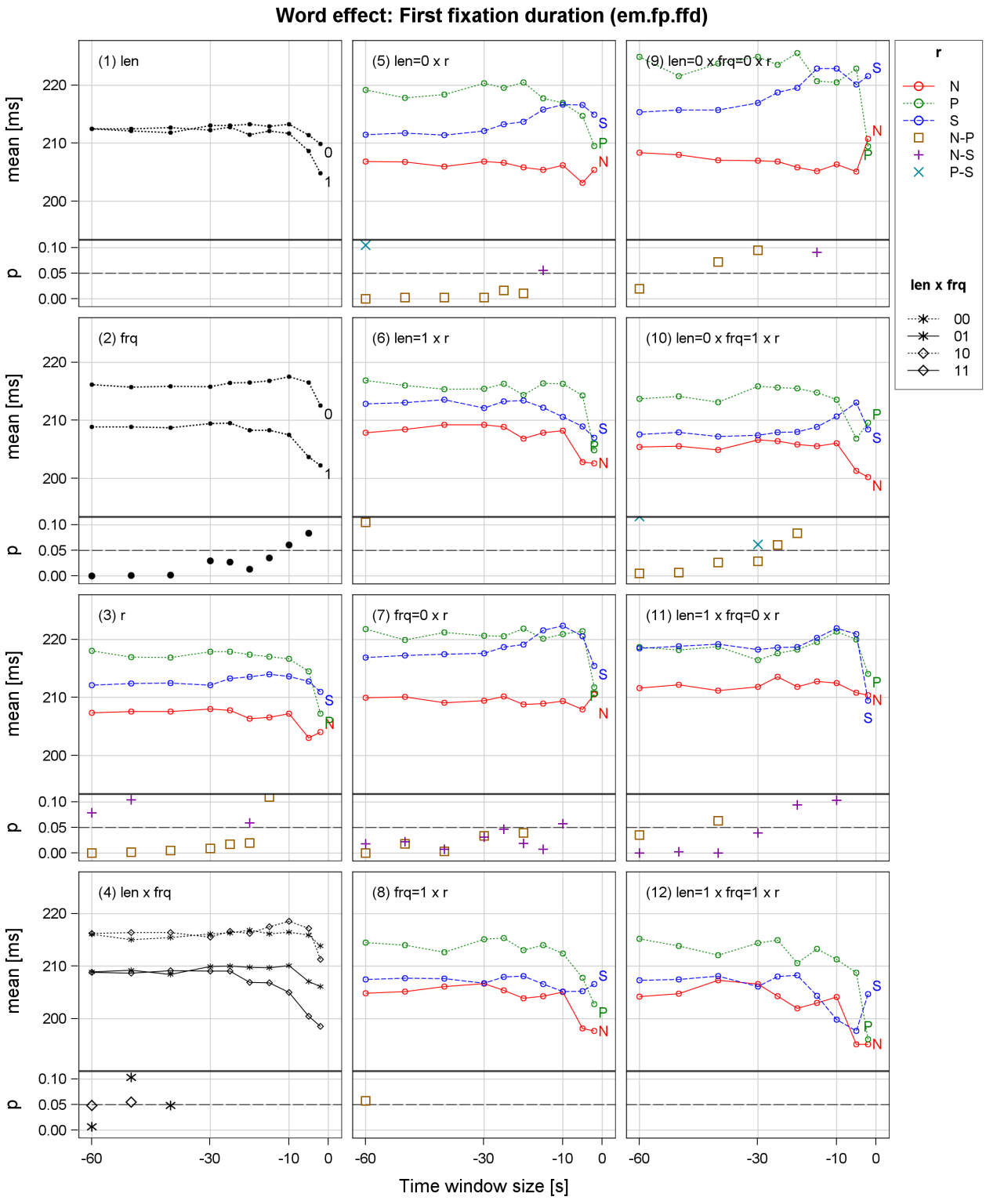


Figure 4.29: Means for main effects and interactions of (1) word length, (2) word frequency, and (3) reading mode (N, P, and S) on first fixation duration ($em.fp.ffd$). Word length and frequency were dichotomized; length was split on four-or-less characters and frequency on its median. Based on fitting model 4.11 (pg. 63) parameterized as shown in Table 4.17 (pg. 64) to a series of data sets sampled through time windows 4–60s wide. Values plotted are predicted population margins; all p-values were adjusted by simulation.

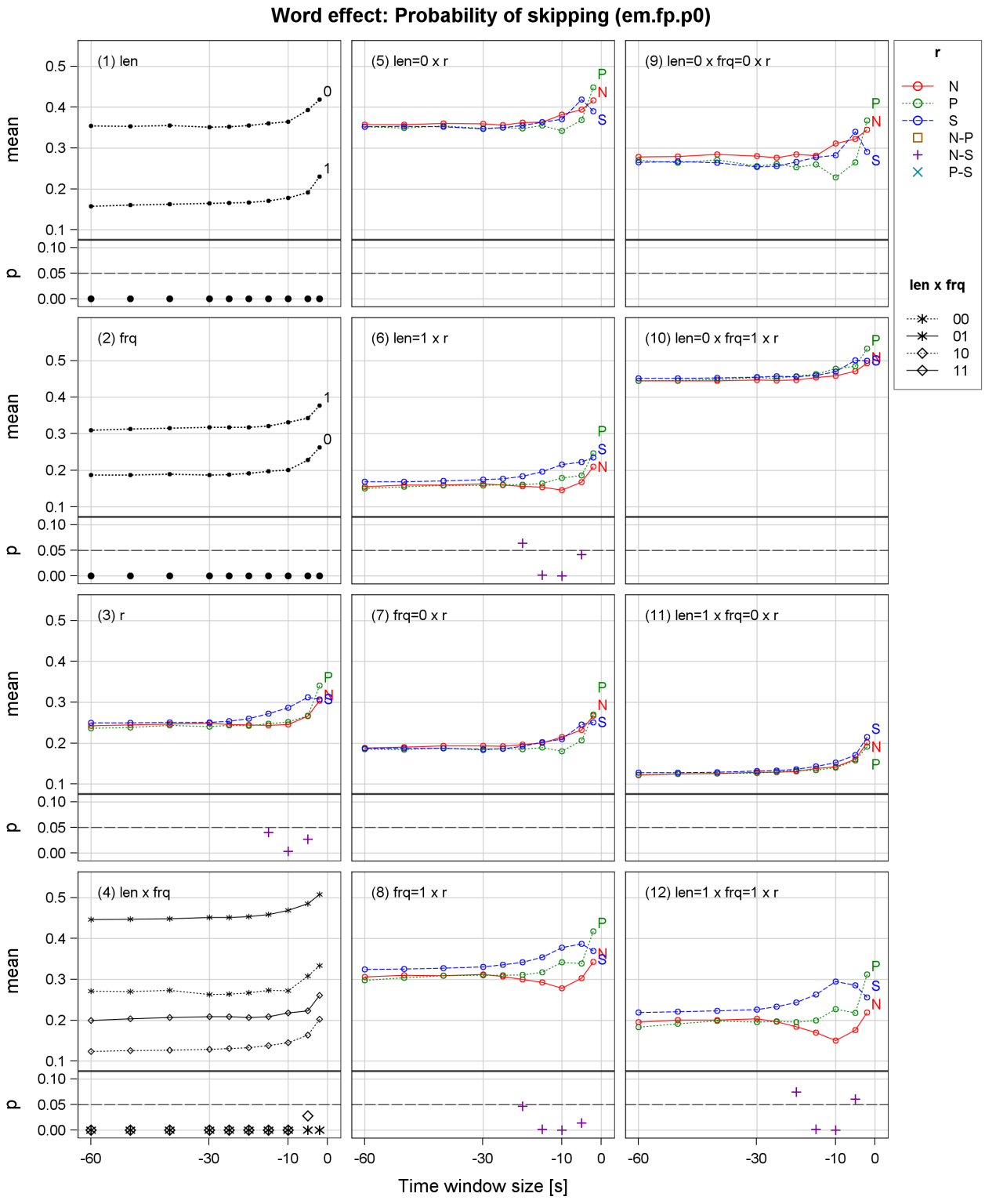


Figure 4.30: Means for main effects and interactions of (1) word length, (2) word frequency, and (3) reading mode (N, P, and S) on probability of skipping (first-pass; $em.fp.p0$). Word length and frequency were dichotomized; length was split on four-or-less characters and frequency on its median. Based on fitting model 4.11 (pg. 63) parameterized as shown in Table 4.17 (pg. 64) to a series of data sets sampled through time windows 4–60s wide. Values plotted are predicted population margins; all p-values were adjusted by simulation.

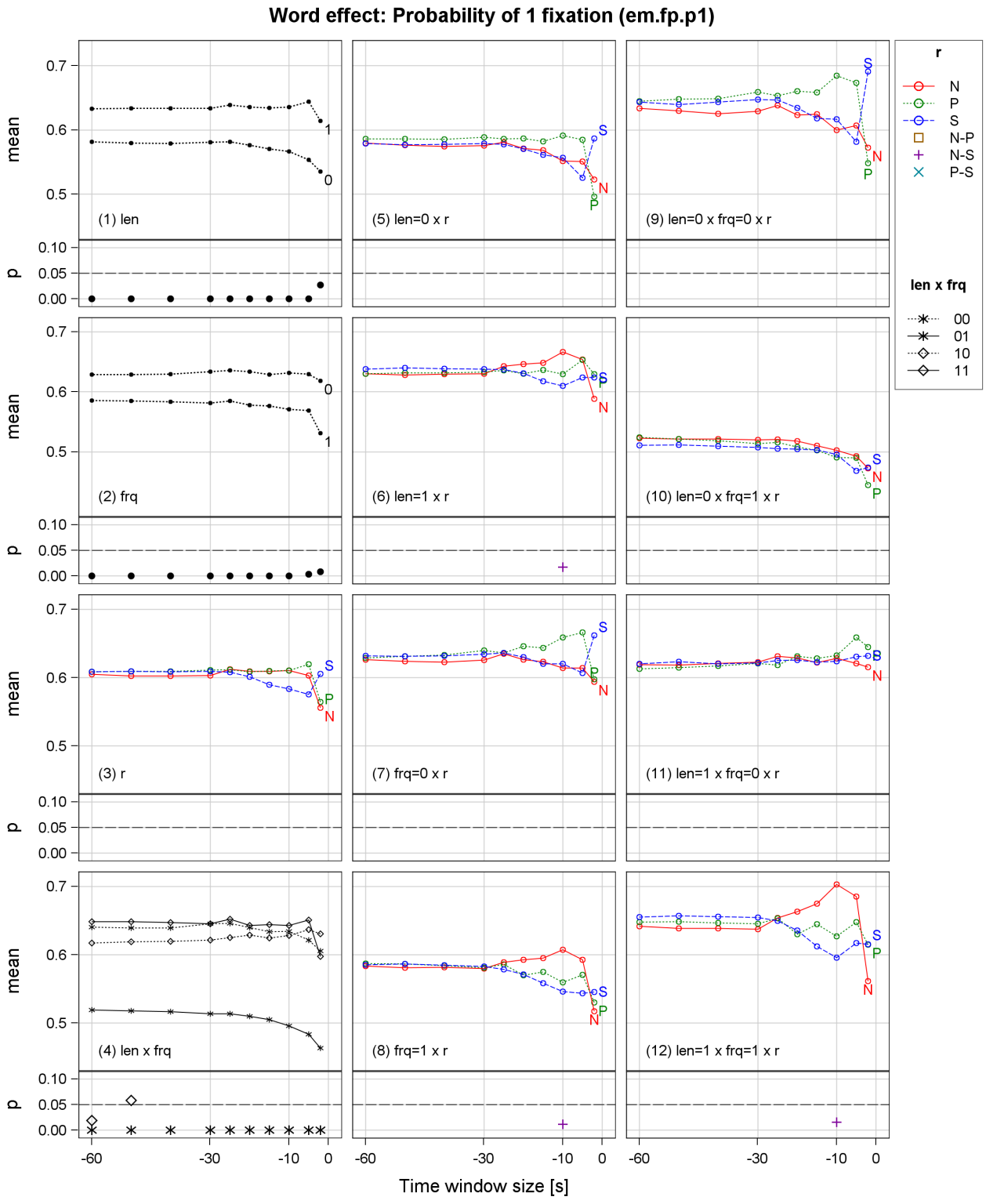


Figure 4.31: Means for main effects and interactions of (1) word length, (2) word frequency, and (3) reading mode (N, P, and S) on probability of one fixation (first-pass; $em.fp.p1$). Word length and frequency were dichotomized; length was split on four-or-less characters and frequency on its median. Based on fitting model 4.11 (pg. 63) parameterized as shown in Table 4.17 (pg. 64) to a series of data sets sampled through time windows 4–60s wide. Values plotted are predicted population margins; all p-values were adjusted by simulation.

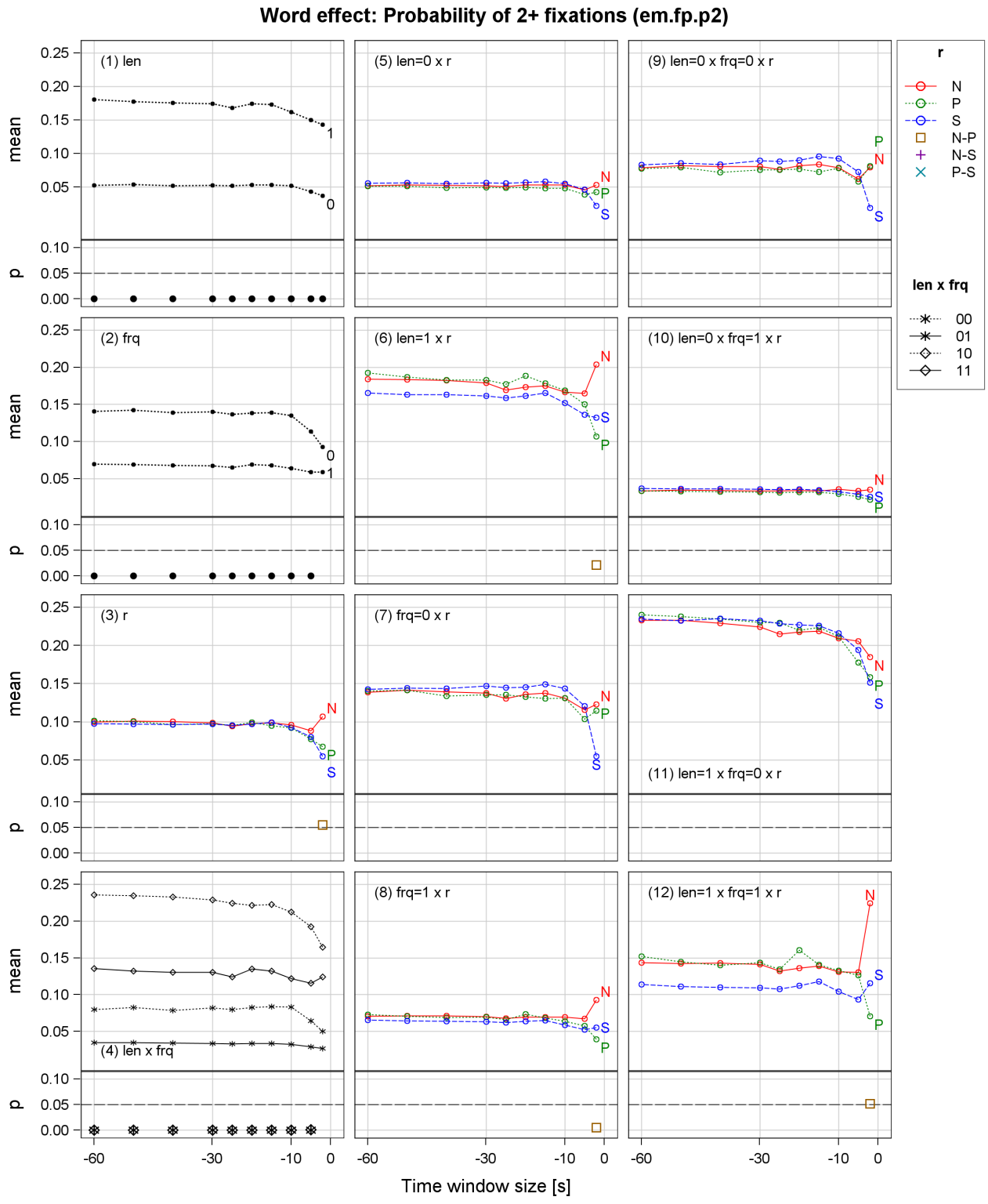


Figure 4.32: Means for main effects and interactions of (1) word length, (2) word frequency, and (3) reading mode (N, P, and S) on probability of more than one fixation (first-pass; $em.fp.p2$). Word length and frequency were dichotomized; length was split on four-or-less characters and frequency on its median. Based on fitting model 4.11 (pg. 63) parameterized as shown in Table 4.17 (pg. 64) to a series of data sets sampled through time windows 4–60s wide. Values plotted are predicted population margins; all p-values were adjusted by simulation.

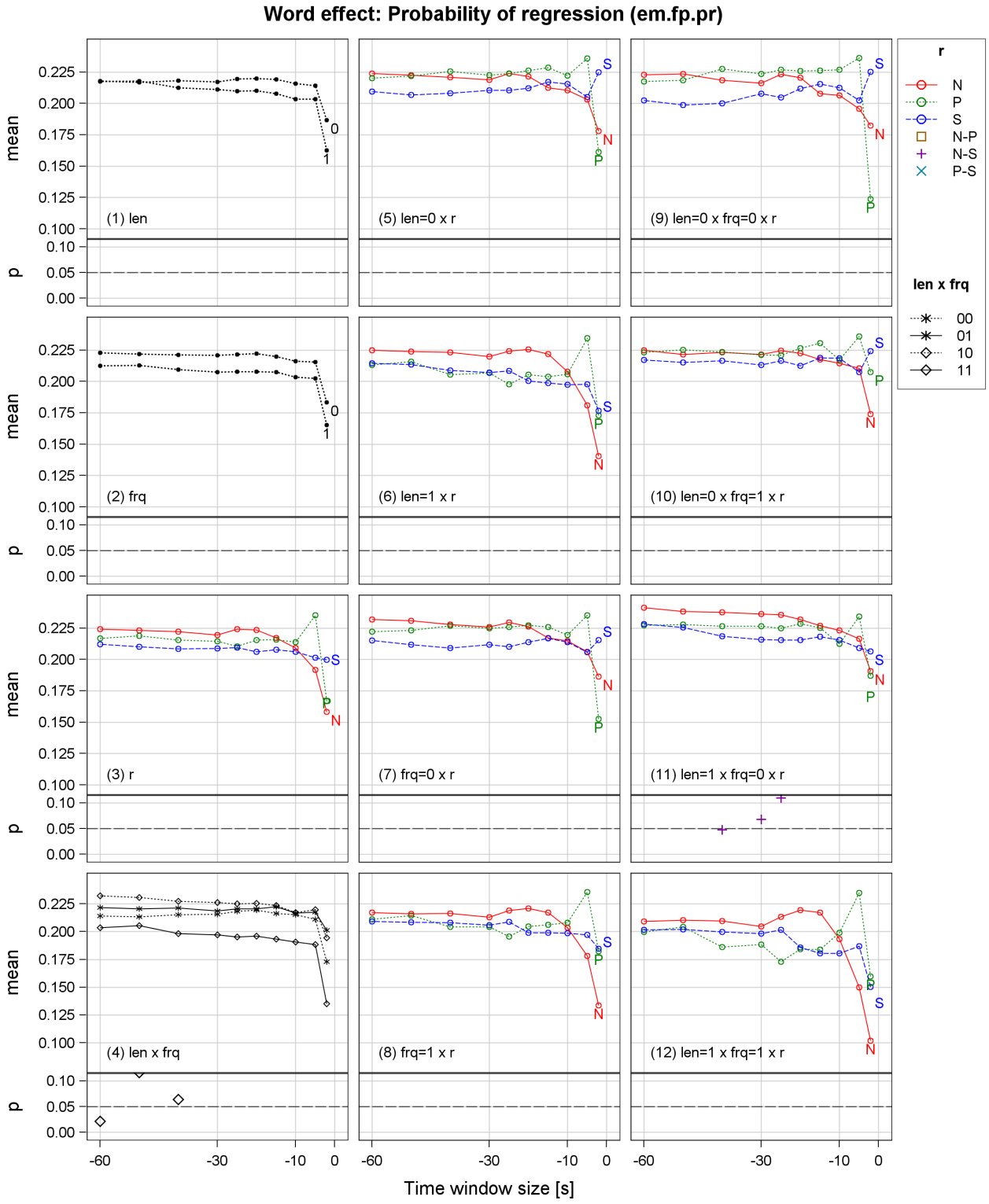


Figure 4.33: Means for main effects and interactions of (1) word length, (2) word frequency, and (3) reading mode (N, P, and S) on probability of regression (first-pass; $em.fp.pr$). Word length and frequency were dichotomized; length was split on four-or-less characters and frequency on its median. Based on fitting model 4.11 (pg. 63) parameterized as shown in Table 4.17 (pg. 64) to a series of data sets sampled through time windows 4–60s wide. Values plotted are predicted population margins; all p-values were adjusted by simulation.

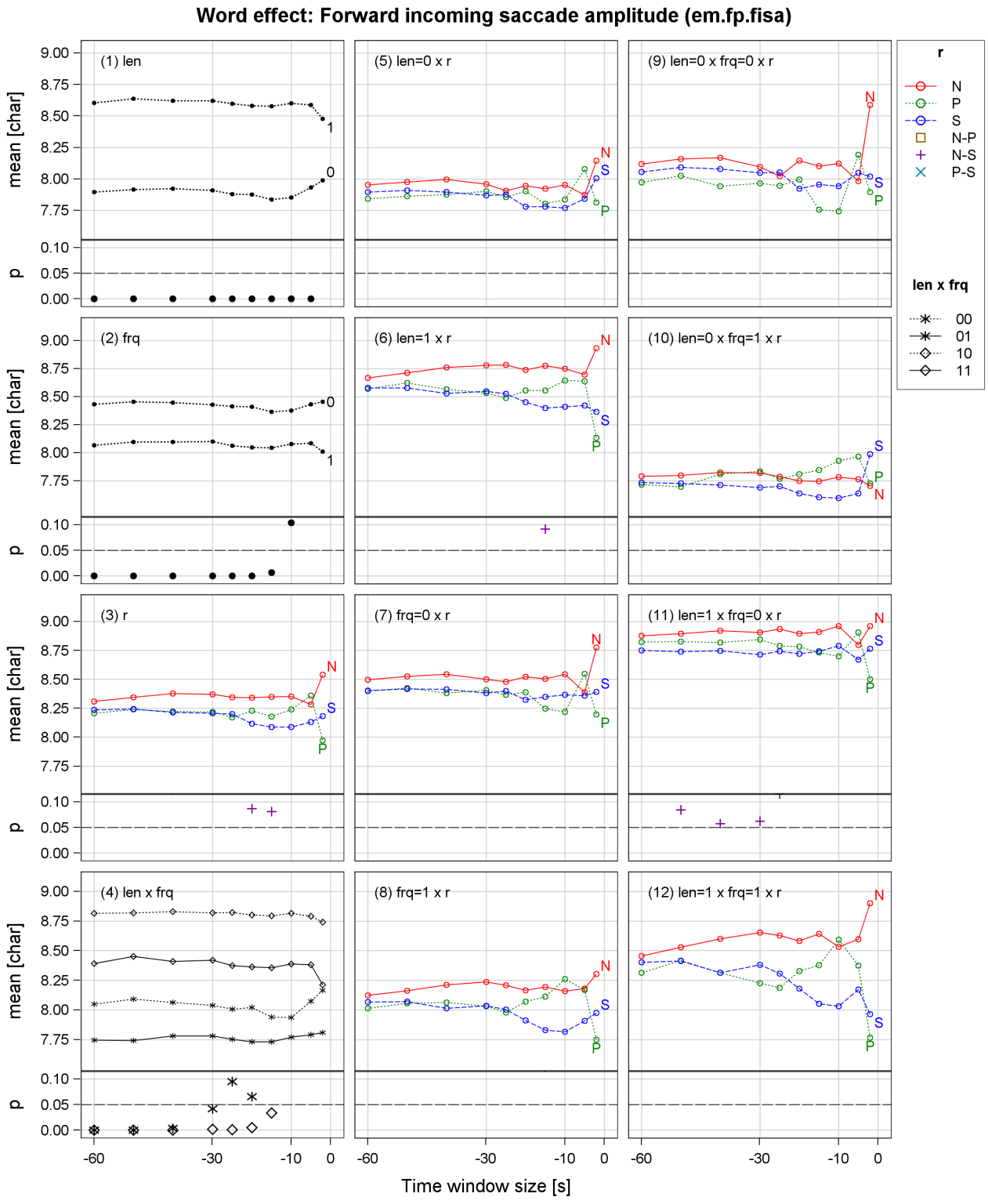


Figure 4.34: Means for main effects and interactions of (1) word length, (2) word frequency, and (3) reading mode (N, P, and S) on forward incoming saccade amplitude (first-pass; $em.fp.fisa$). Word length and frequency were dichotomized; length was split on four-or-less characters and frequency on its median. Based on fitting model 4.11 (pg. 63) parameterized as shown in Table 4.17 (pg. 64) to a series of data sets sampled through time windows 4–60s wide. Values plotted are predicted population margins; all p-values were adjusted by simulation.

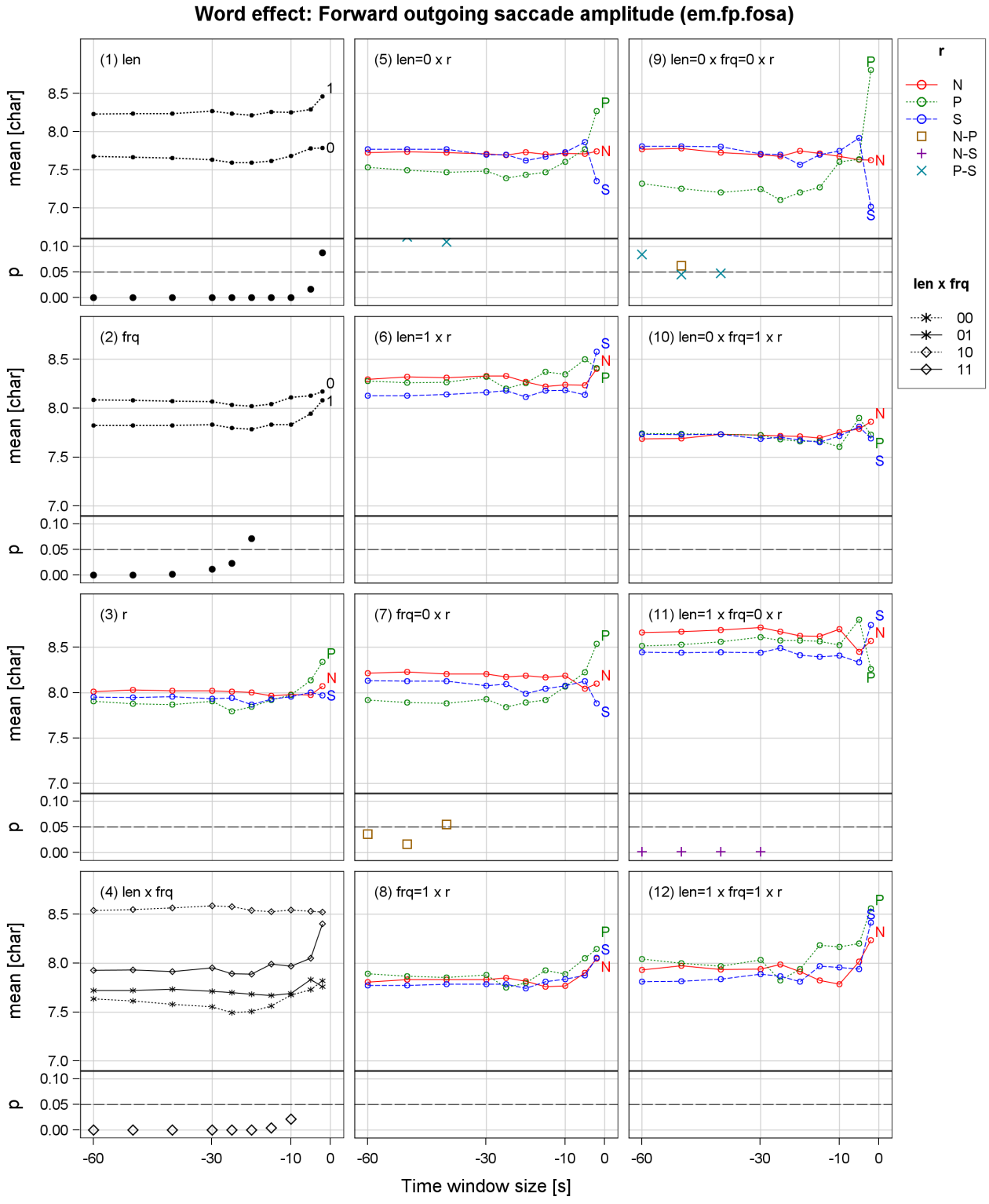


Figure 4.35: Means for main effects and interactions of (1) word length, (2) word frequency, and (3) reading mode (N, P, and S) on forward outgoing saccade amplitude (first-pass; $em.fp.fosa$). Word length and frequency were dichotomized; length was split on four-or-less characters and frequency on its median. Based on fitting model 4.11 (pg. 63) parameterized as shown in Table 4.17 (pg. 64) to a series of data sets sampled through time windows 4–60s wide. Values plotted are predicted population margins; all p-values were adjusted by simulation.

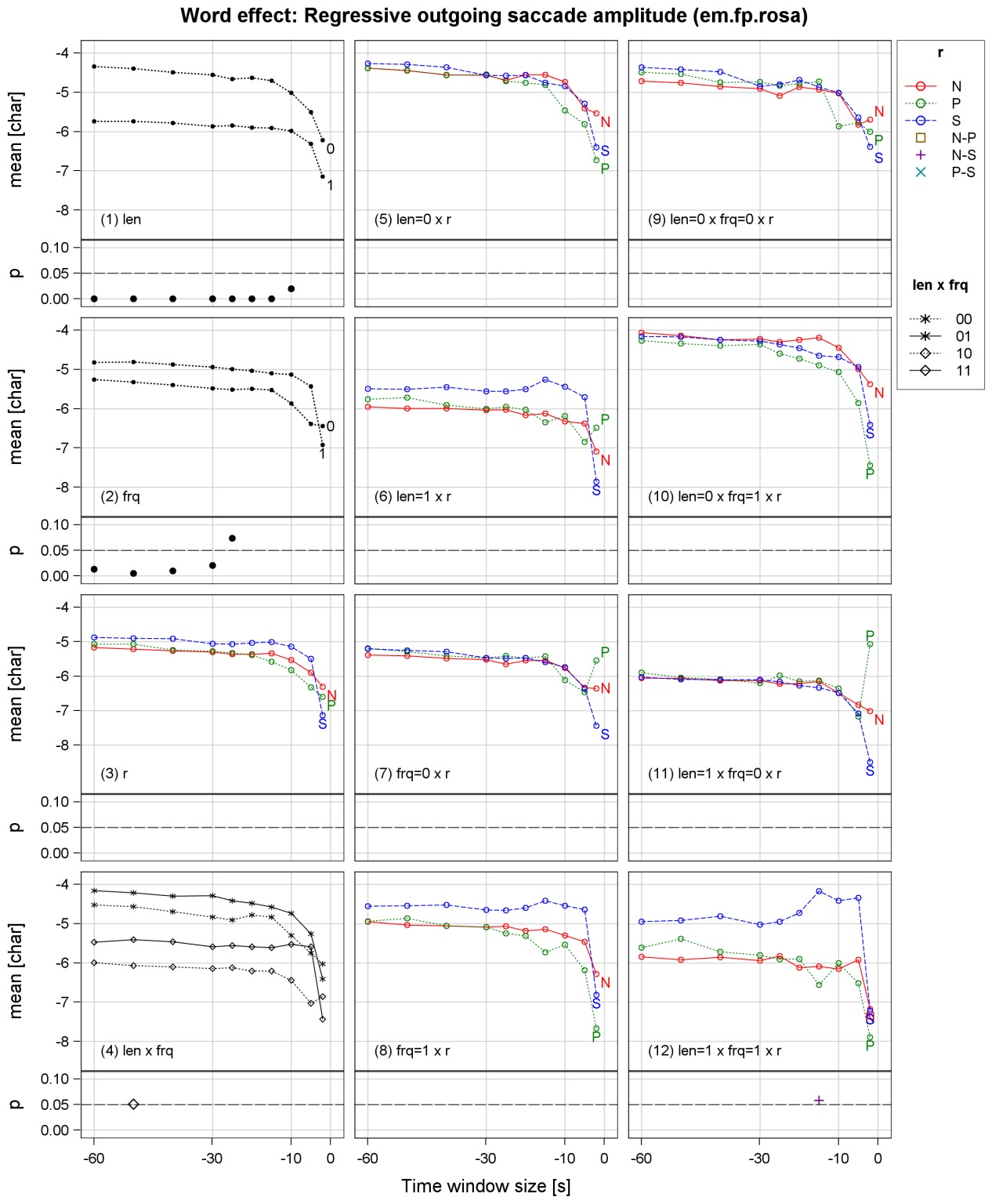


Figure 4.36: Means for main effects and interactions of (1) word length, (2) word frequency, and (3) reading mode (N, P, and S) on regressive outgoing saccade amplitude (first-pass; $em.fp.rosa$). Word length and frequency were dichotomized; length was split on four-or-less characters and frequency on its median. Based on fitting model 4.11 (pg. 63) parameterized as shown in Table 4.17 (pg. 64) to a series of data sets sampled through time windows 4–60s wide. Values plotted are predicted population margins; all p-values were adjusted by simulation.

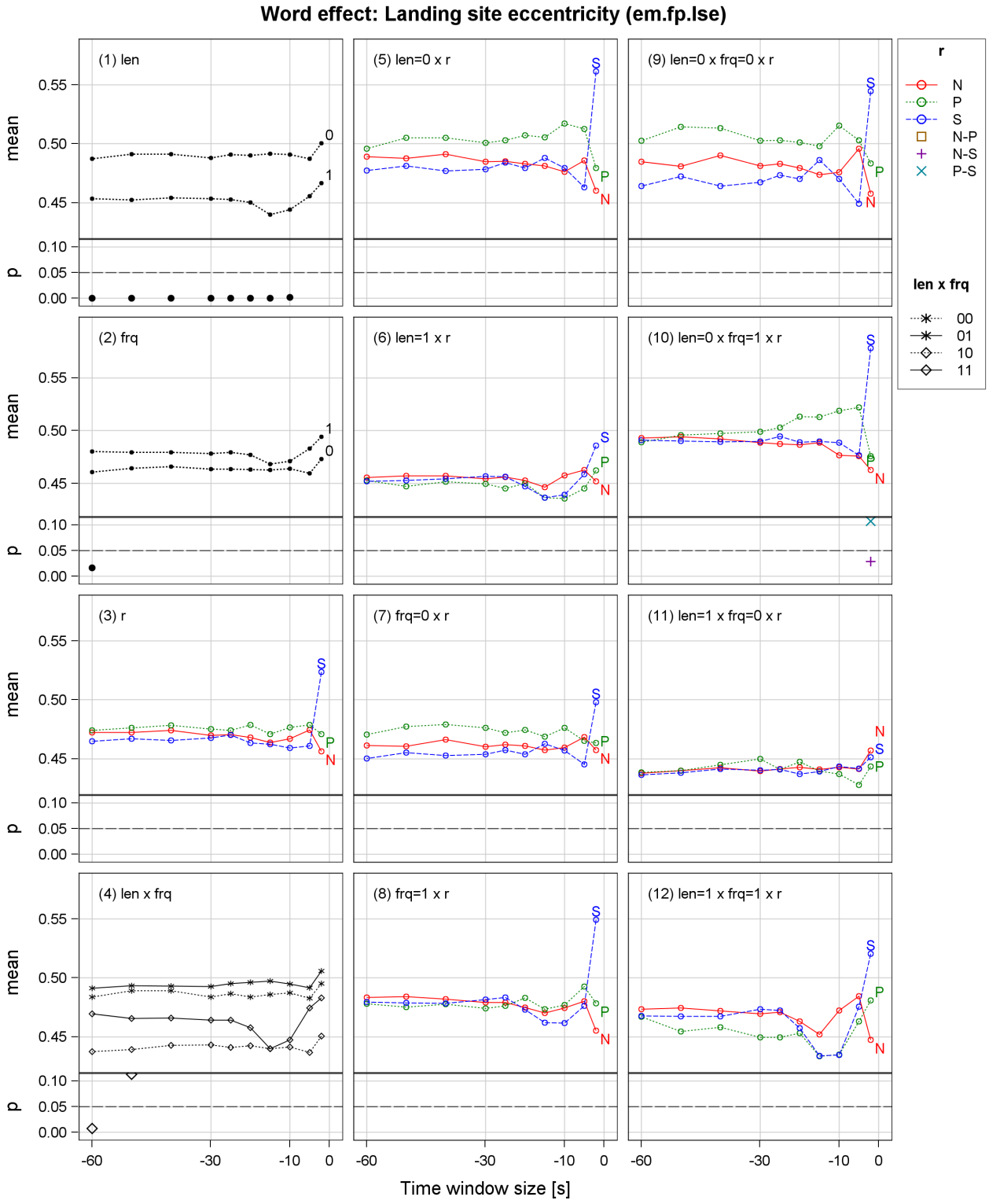


Figure 4.37: Means for main effects and interactions of (1) word length, (2) word frequency, and (3) reading mode (N, P, and S) on landing site eccentricity (first-pass; $em.fp.lse$). Word length and frequency were dichotomized; length was split on four-or-less characters and frequency on its median. Based on fitting model 4.11 (pg. 63) parameterized as shown in Table 4.17 (pg. 64) to a series of data sets sampled through time windows 4–60s wide. Values plotted are predicted population margins; all p-values were adjusted by simulation.

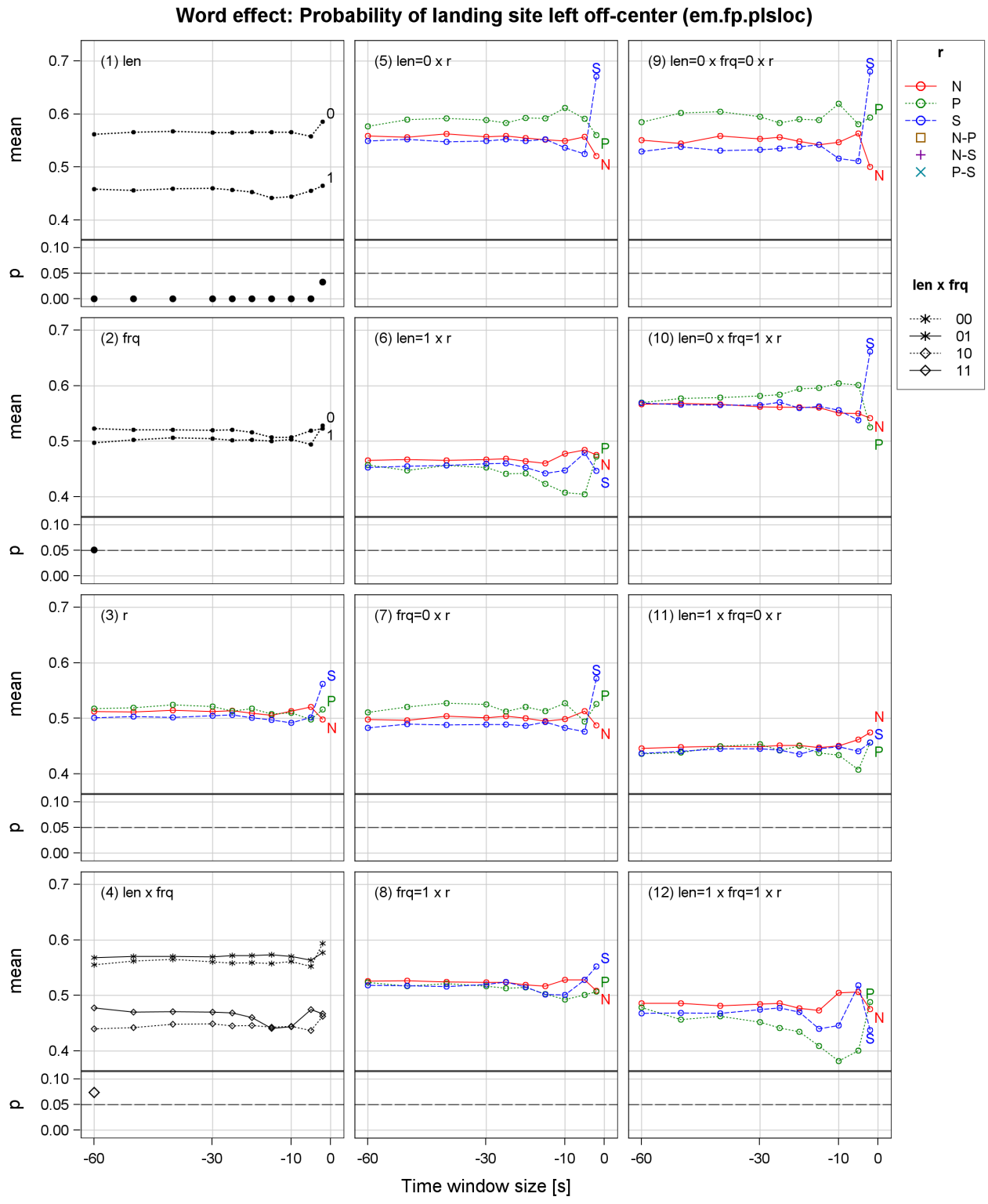


Figure 4.38: Means for main effects and interactions of (1) word length, (2) word frequency, and (3) reading mode (N, P, and S) on probability of landing site left-off-center (first-pass; $em.fp.plsloc$). Word length and frequency were dichotomized; length was split on four-or-less characters and frequency on its median. Based on fitting model 4.11 (pg. 63) parameterized as shown in Table 4.17 (pg. 64) to a series of data sets sampled through time windows 4–60s wide. Values plotted are predicted population margins; all p-values were adjusted by simulation.

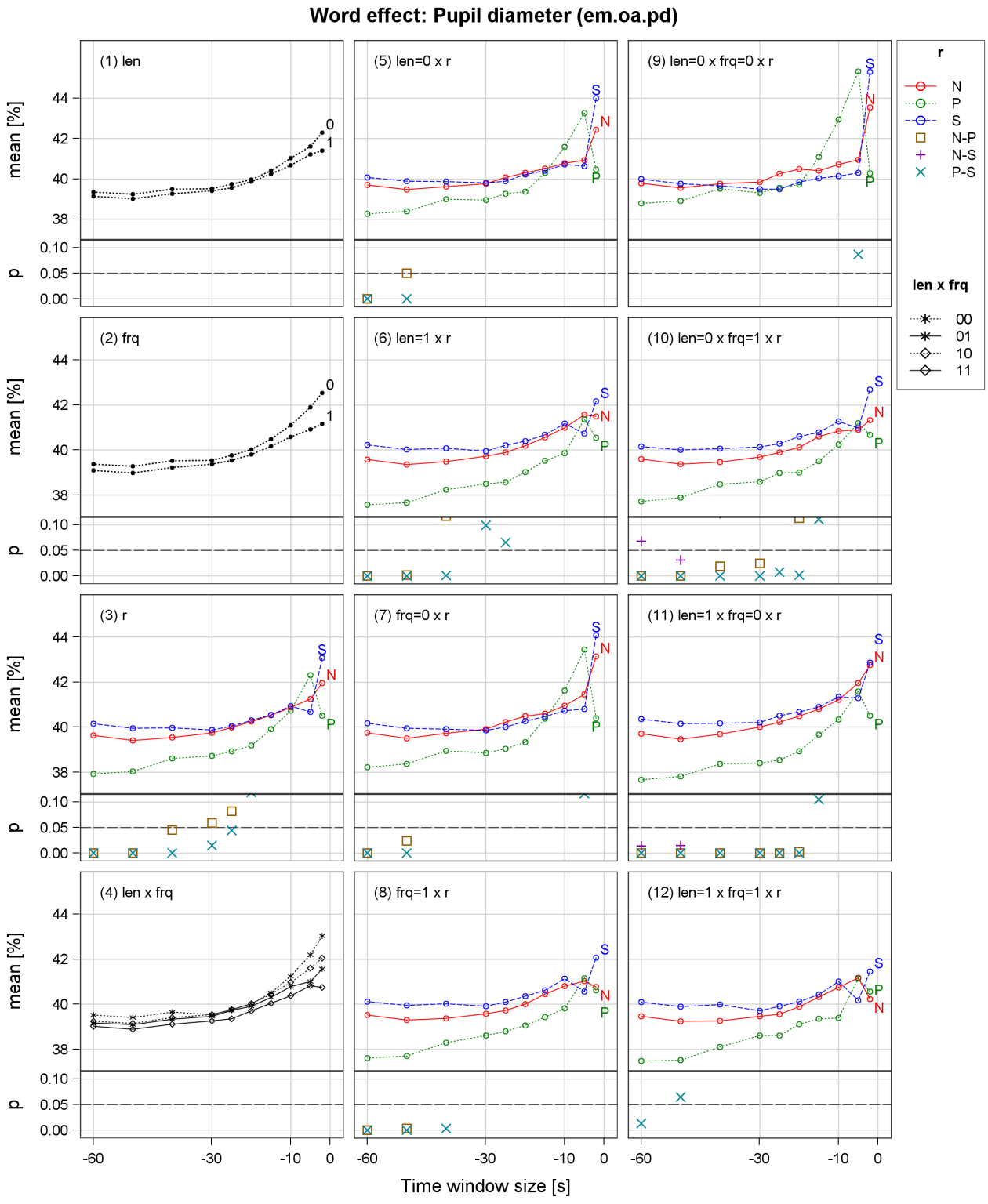


Figure 4.39: Means for main effects and interactions of (1) word length, (2) word frequency, and (3) reading mode (N, P, and S) on pupil diameter (first-pass; *em.oa.pd*). Word length and frequency were dichotomized; length was split on four-or-less characters and frequency on its median. Based on fitting model 4.11 (pg. 63) parameterized as shown in Table 4.17 (pg. 64) to a series of data sets sampled through time windows 4–60s wide. Values plotted are predicted population margins; all *p*-values were adjusted by simulation.

Chapter 5

Detecting Mindless Reading

In Chapter 4, I have investigated differences between normal and mindless reading that can be found in the eye-movement reading data I collected and which subject-level and eye-movement variables could prove useful in discriminating between these two attentive states. In this chapter, I use all these results as a starting point in my endeavors to develop statistical models for mindless reading identification.

I conducted all analyses I report in this chapter in R (R, 2013).

5.0.1 Prior Research

The problem of discovering mind-wandering in readers, while recognized as being important (Smallwood, 2011), has attracted limited attention so far. Schad, Nuthmann, & Engbert (2012) used gaze duration on 14 short and long words fixated prior to mind-wandering onset to predict mindless reading using the logistic regression model. Unfortunately, although the results reported are promising, they are also inconclusive in that only classification accuracy is mentioned and no discussion of issues like class imbalance is provided. Furthermore, it is not clear what evaluation scheme was employed in that research. For example, if resubstitution validation was used then even good performance may turn pale when a classifier is actually deployed and used on data it has not been trained on.

D'Mello, Cobian, & Hunter (2013) used a battery of 33 classification algorithms to detect probe-caught mindless reading. While insufficient detail about that research is available at this point, the authors report (using the kappas measure of agreement and leave-several-subjects-out validation) promising results obtained after correcting for class imbalance. Surprisingly though, while it is not clear exact which variables are used, text-unrelated variables are found to be more important than text-related ones.

5.0.2 Failed Early Attempts

During my initial attempts to detect mindless reading, I spent a considerable amount of time on estimating and evaluating several generations of discrete Bayesian classification models. I tried several network structures and paired them with many data sets. More specifically, I tried several time window sizes and several discretization schemes. As it turned out, all that was to no avail. I have eventually concluded that my lack of success had to do with discretization. That is, the differences between distributions of continuous variables I was interested in for normal and mindless reading were small, and discretizing them lead to "smearing" these differences away by binning together values which should be different to start with. In the end, I had little choice but to abandon discrete models altogether and use continuous (and hybrid) models to completely sidestep discretization. I choose to keep the details of the Bayesian models I tried compartmentalized for the time being.

5.1 Experimental Setup

After a complete lack of success with a range of discrete Bayesian networks, I decided to use one of the simplest classification models able to handle continuous variables: The logistic regression model. Because it is simple it is also one of the most widely understood and utilized statistical models and therefore it makes for as a good baseline for any future investigations.

5.1.1 Experimental Reading Data

The current experiment resulted in collecting what arguably is the largest mindless reading data set obtained to date. By its design, this experiment is in-line with most reading experiments which focus on recruiting a large number of readers each of which reads for a typical duration of about an hour. However, longer experiments have been conducted as well. For example, Reichle, Reineberg, & Schooler (2010) run an insightful experiment with a

Table 5.1: Similarities and differences between the Reichle, Reineberg, & Schooler (2010) and the current (2014) experiments. Means with SDs in brackets where appropriate.

	2010	2014
Subjects retained	4	112
Sessions	14	1
Session time	1h	2h
Working mem. task	No	Yes
Total reading time	13h 30m (39m)	71m 43 s (10m 9s)
Pages read	377 (0)	49.77 (14.05)
Calibration every	1 chapter	2-3 chapters
Text compreh.	0.81 (0.03)	0.68 (0.17)
Probe-caught ratio	0.09	0.24
Text used	*Sense and Sensibility*	
Eye-tracker used	EyeLink 1000	
Probes every	2-4 min	
Self-reports	Yes	

slightly different design. Namely, instead of using a large number of readers they opted to use few subjects and ask them to read for much longer, roughly 14 hours, one hour per experimental session. It is entirely possible that their design gave rise to somewhat different mind-wandering behavior than the one I had captured in my experiment. For example, it is plausible that once familiar with the eye-tracker, subjects participating in the several-session-long experiment mind-wander more "naturally" due to the fading effect of novelty of the situation. Having access to data from both experiments, I decided to fit all my classification models to both. For brevity, in the remainder of this chapter I refer to the experiment conducted by Reichle, Reineberg, & Schooler (2010) and to my experiment simply as 2010 and 2014, respectively (same goes for the data collected through these experiments). Table 5.1 compares the two experimental designs and lists several vital statistics.

5.1.2 Data Sets

Because the exact duration of a mindless reading episode is not known, I estimated and evaluated all my classifiers on a varying amounts of eye-movement data, ones that were encompassed by time windows of size 2–30s all culminating in a probe or a self-report. That allowed me to asses the feasibility of classifiers having three different *reaction speeds*: *fast* (2 and 5 seconds), *medium* (10 seconds), and *slow* (20 and 30 seconds). For example, a fast classifier will give its prediction about whether or not a reader is reading mindlessly by looking at 2 or 5 seconds of word-level and eye-movement data, depending on time window. Naturally, faster classifiers are preferred, but because they look at a smaller portion of the data their performance may suffer as a result.

Additionally to varying time window size, I also varied the minimum length of a words which classifiers considered. That was operationalized by including only words longer than a designated threshold in the estimation and evaluation data sets. I used six thresholds (2, 4, 6, 8, 10, and 12 character spaces) and also allowed the word length to be unconstrained (i.e., minimum word length of 0).

The way I varied time window size and word length was by creating a series of data sets with the possible combinations of these two variables and subsequently using these data sets for model estimation and evaluation. I did not create any data sets based on word frequency and included it in each candidate model as I describe later.

5.1.3 Classification Tasks

Each reader which participated in either the 2010 or the 2014 experiment at any point in time could be in one of three reading modes: Normal (or mindful; N), mindless probe-caught (i.e., plausibly not being meta-aware; P), and mindless self-caught (i.e., which eventually led to them regaining meta-awareness; S). It is likely that transitions between these modes are quite fluid (Schad, Nuthmann, & Engbert, 2012), but here I treat them as discrete and mutually exclusive states a reader can be in. In the context of classification, states like these are typically referred to as *classes*.

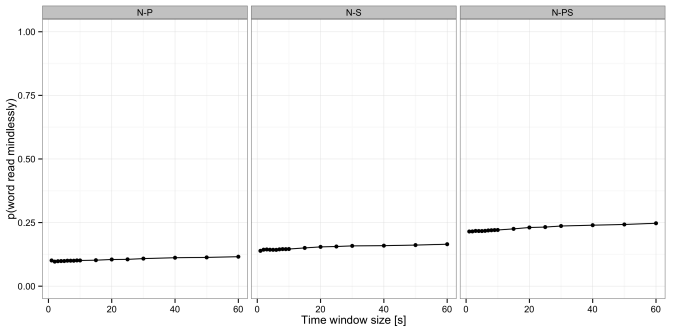
Instead of attempting to forge three-class classifiers (i.e., N-P-S), I chose to cast this problem as three separate binary classification tasks (i.e., N-P, N-S, and N-PS). The reason for this is two-fold. First, as the simplest and more common case, binary classification has been extensively studied and as a result many standard evaluation approaches are available for it (Section 5.1.5). Second, and more importantly, it is my impression that when deployed in an actual reading experiment classifiers I am attempting to develop would also be expected to perform binary classification. More specificaly, their job would be to either (a) distinguish between mindful and mindless probe-caught reading (i.e., P-N; the reader could still be asked to self-report mindless reading) or (b) distinguish between mindful and mindless reading of any kind (i.e., N-PS; the reader would no longer be asked to self-report mindless reading).

These three distinct binary classification tasks map onto the three types of experiments I foresee made possible by research like this current one. The first type of an experiment (i.e., N-P) would study mindless reading in its "purest" form, i.e., study probe-caught mindless reading but with probe-caught ratios (i.e., the proportion of probe actually discovering mindless reading) higher than those expected to occur if a random probes were utilized. The reason why probe-caught mindless reading should represent instances of pure mind-wandering is that read-

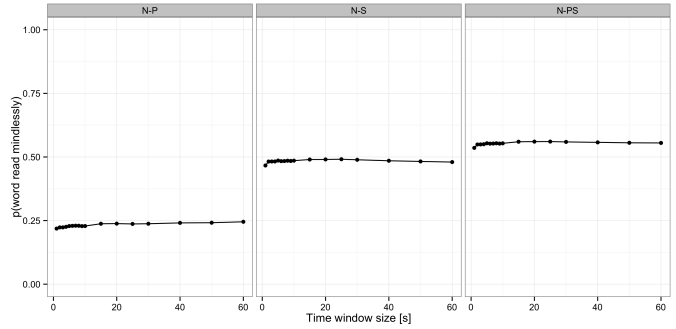
ers are actually caught mind-wandering and therefore it can be assumed that on at least a portion of probe hits they are not aware (or not aware enough) of their mind-wandering.

The second binary classification task (i.e., N-S) would be useful in the study of issues surrounding consciousness and meta-consciousness. That is because mindless reading should typically be self-reported right after a person regains their meta-cognitive abilities which allow them to inspect the content of their consciousness.

Finally, the third type of an experiment (i.e., N-PS) would study mindless reading through a gaze-contingent paradigm (see, e.g., Rayner & Bertera, 1979) where changes to the stimulus would still be constrained to the immediate location of current fixation but were to occur only when the reader was mindless. In such an experiment, subjects would not be probed or asked to self-report mindless reading and the emphasis would be on their reactions to stimulus manipulation. Incidentally, such an experimental setup would also help to avoid an important source of contamination of the cognitive processes active during reading. Namely, the contamination introduced by inducing higher levels of self-awareness and introspection through the instruction to self-report mindless reading.



(a) 2010



(b) 2014

Figure 5.1: Probability that a word was read mindlessly in the two experiments I used: Reichle, Reineberg, & Schooler (2010) and the current one (2014). Class imbalance apparent especially in the 2010 data.

5.1.4 Class Imbalance

Figure 5.1 shows the probability that a word was read mindlessly in both the 2010 and 2014 experiments as a function of time window size and classification task. Class imbalance is apparent across the board in the 2010 data with less than a quarter of words contributing observations to the mindless reading class. The 2014 data is less afflicted by this problem which is manifested only in the context of the N-P classification task. This difference between these two studies could be due to two factors. First, sessions were longer in the 2014 experiment (2h compared to 1h). Second, the 2014 experiment opened with complex working memory span task. Both of these factors could have made subjects more fatigued and consequently more prone to lapsing into mindless reading.

If the goal of a binary classifier is to detect the minority class (as is in the current research), class imbalance is a problem because the more severe imbalance the harder it is for the classifier to consider the under-represented class (Kubat & Matwin, 1997; Japkowicz & Stephen, 2002). Additionally, the scarcity of one of the classes may lead to poor estimation of a model's classification performance. For example, a model may seem to be performing well while in reality it is simply favors the majority class.

There are two types of approaches that aim at mitigating this problem: Solutions at the learning level and solutions at the data level (Kotsiantis et al., 2006). Learning-level solutions center around using algorithms that take class imbalance into account directly. This often means developing new estimation algorithms tailored specifically to the data at hand. Data-level solutions on the other hand focus on modifying class distribution in the estimation data set before model estimation commences.

For the purpose of this dissertation, I used data level solutions to correct for the class distribution skewness. I did that by using all the majority class observations and oversampling the minority class observations until the proportion of observations from that class reached approximately 50%. Because data sets I used contained many observations per subject, sampling at the level of individual observation is incorrect. Instead, I used stratified sampling with subjects being the strata. That is, I effectively injected new subjects by duplicating some of the existing ones. Because different subjects have different number of observations for the rare class and they are resampled at random achieving the exact 50-50 class proportion was unlikely to happen; instead I used a nearly-50-50 class proportion.

5.1.5 Estimation and Evaluation

I estimated all GLM models with the *glm()* function from the "stats" package which uses iteratively reweighted least squares (IWLS) for model fitting. The GLMM models were estimated with the *glmer()* function from the "lme4" package. I used Laplace approximation (or adaptive Gauss-Hermite quadrature with one point per axis) for model fitting.

To evaluate the estimated models, I used four validation methods: (1) Resubstitution, (2) bootstrapped resubstitution, (3) 10-fold bootstrapped cross-validation, and (4) 10-fold bootstrapped leave-subjects-out cross-validation. While the first three methods are quite straightforward (I explain them in the next subsection), in the fourth one I simply ensured that the training and test data sets did not contain the same subjects. By doing that I simulated a desirable deployment scenario in which a classifier is used on eye-movement data from a reader it has not seen before. In other words, if that fourth evaluation method indicates poor classification performance then a classifier cannot be used to reliably classify new readers. Now, I shed some light on the different evaluation methods.

5.1.6 Evaluation: Cross-validation

Cross-validation (*CV*) is a resampling method for validating a predictive model by estimating its generalization error (Geisser, 1993; Plutowski, Sakata, & White, 1994; Weiss & Kulikowski, 1991). It is better than methods based on residuals because it tells us something about future behavior of a model, i.e., its behavior given data that has not yet been observed. CV guards against overfitting present in the case of *resubstitution validation* (which uses the same data for both estimation and evaluation). It is also less wasteful of data and more stable than hold-out validation.

CV comes in several flavors, but each is based on the idea of splitting the data into the estimation (or training) and evaluation (or test) sets. Random subsampling CV (RSCV) performs K splits, each randomly selecting a fixed number (K) of observations without replacement. K-fold CV (KFCV) on the other hand partitions the data into K parts. At every of the K validation steps, a different part is used as the estimation set. The advantage of KFCV over RSCV is that all observations in the data set are used for both model estimation and model evaluation. Data can be stratified for KFCV to ensure equal proportions of values of the stratification variable (or variables) are present in both sets at every fold. Repeated KFCV is also possible. The true error for both RSCV and KFCV is estimated as the average error rate

$$\epsilon = \frac{1}{K}\sum_{i=1}^{K}\epsilon_i \ .$$

Leave-one-out CV (LOOCV) is an extreme case of KFCV in that it divides the data set into as many parts as observations present in the data set. At each validation step, the one observation that was set aside is used to calculate the prediction error of a model. In other words, the evaluation data set is of length 1 and the estimation data set is of length $n-1$ and there are $n-1$ estimation–evaluation runs. In LOOCV, the true error is given as

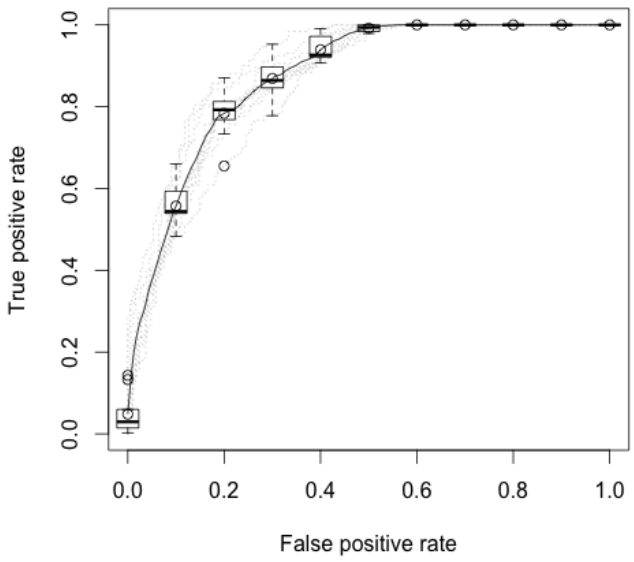
$$\epsilon = \frac{1}{N}\sum_{i=1}^{N}\epsilon_i \ .$$

The optimal number of folds depends on the size of the data set and the purpose of a model. For large data sets, even few (e.g., $K = 3$) folds can be accurate. For small data sets, it is important to use as many observations for training as possible. LOOCV is preferred for estimating generalization error for continuous functions (regression), while KFCV for discontinuous functions (classification; the task I focus on in this dissertation). LOOCV is known to be almost unbiased but also to have high variance, leading to unreliable estimates (Efron, 1983). A common choice for KFCV is $K = 10$ (which I also use in this research).

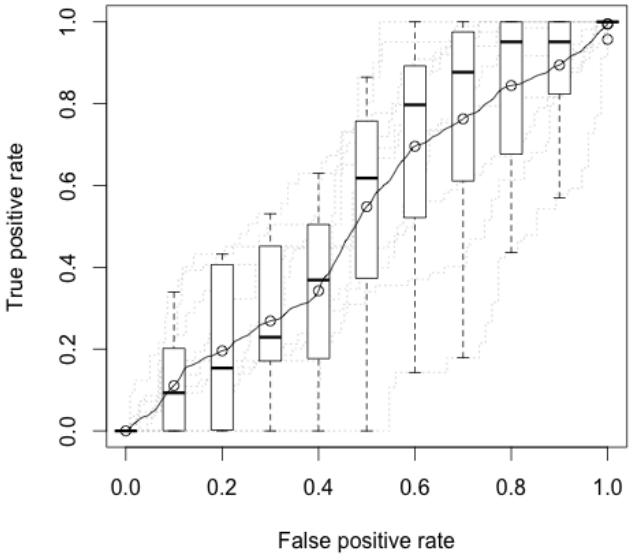
5.1.7 Evaluation: ROC and ROC AUC

The *receiver operating characteristic* (*ROC*; or simply *ROC curve*) is a graphical illustration of a binary classifier's performance as its discrimination threshold is varied. The ROC curve is constructed by plotting the true positive rate (i.e., the proportion of true positives out of actual positives; TPR; sensitivity) against the false positive rate (i.e., the proportion of false positives out of actual negatives; FPR; specificity) for various values of the discrimination threshold. Figure 5.2 shows two examples of ROC curves (both are actual curves I obtained). The top example shows a very good classification performance as indicated by the curve being very close to the top-left corner of the graph. On the other hand, the bottom example shows an essentially chance-level classification performance (i.e., one which would result from performing the classification with a fair coin toss). By varying the discrimination threshold we move along the curve. By picking any particular point on the curve we determine the TPR and FPR of our classifier. Visible especially on the top example is the natural tradeoff between the TPR and FPR. For example, increasing the TPR happens at the cost of also increasing the FPR. What levels of both of these rates are acceptable depends on the particular classification task, but good classifiers will always have the ROC curve close to the top-left corner.

While the ROC curve is a very useful measure of classification performance it is not well suited for comparison



(a) Good classification performance (above chance-level)



(b) Bad classification performance (essentially chance-level)

Figure 5.2: Examples of ROC curves.

of multiple classifiers, a problem I am facing in this research. One way to use the ROC curve in this situation is to summarize it with one number. The typical choice for such a summary is the *area under the ROC curve* (or *ROC AUC* for *Area Under Curve*), a number between 0 and 1. It is clear that the two example ROC curves from Figure 5.2 have very different AUCs. Specifically, while the ROC AUC for the bottom example is roughly 0.5 (because the curve splits the plot in more-or-less two equal parts), the ROC AUC for the top example will be larger then 0.5 and equal to about 0.75. A theoretical perfect classifier will have the ROC AUC of 1. I use the ROC AUC as a one-number measure of classification performance in all result plots I report in the remainder of this dissertation.

Note also that the two example ROC curves plotted using solid line are actually averages of 10 cross-validation runs. The ROC curves for the 10 individual cross-validation folds are plotted using dashed line. This illustrates one of the strengths of cross-validation. Namely, one of two exceptionally good or bad ROC curves do not influence the average classification performance in an arbitrary way.

5.1.8 Candidate Models

In my experiment I investigated logistic regression models from two families: Generalized linear models (GLMs) and generalized linear mixed models (GLMMs). Naturally, the difference between them is the presence of a random effect term(s) in a GLMM model; I used subject random effect $s_i \sim N(0, \sigma_s^2)$. Table 5.2 lists model templates which I used to generate the full list of models. In that table, $em.var$ is a pseudo-variable defined as a vector of actual eye-movement variables

$$em.var = \langle em.ffd, em.gd, em.nb, em.nfpf, em.nfpr, em.pd, em.sfd \rangle.$$

By this I mean to say that each model template containing the $em.var$ pseudo-variable "unrolls" into seven actual models, each containing one actual eye-movement variable. Consequently, my templates generated 73 GLM–GLMM model pairs or a total of 146 models.

As will become evident later, I used all 146 models with the 2014 data but only nine of them on the 2010 data. That is because no subject-level variables were collected during the 2010 experiment. Table 5.2 specifies which model templates were used with the 2010 data. Additionally, IDs of model templates listed in the first column of that table are referenced in result figures I report in the next section.

As far as subject-level variables are concerned, I focused on those which have been found (by other researchers or me) to impact mind-wandering behavior. I list all eye-movement and subject-level variables in Table 5.3.

Table 5.2: List of model templates I investigated in my modeling experiment. "N vars" denotes the number of eye-movement and subject variables included in each "unrolled" model. "N models" denotes the number of actual models each pseudo-formula "unrolls" into for the 2010 and the 2014 data. That number is seven for all model templates containing the pseudo-variable $em.var$ (see main text for explanation) and is zero for some templates for the 2010 data because no subject-level variables were collected in that experiment. π denotes the probability of the reader mindlessly reading (P, S, or P+S, depending on the data set used). $w.len$ and $w.frq$ denote word length and frequency, respectively. $(all.em.vars)$ and $(all.sub.vars)$ are placeholders for all eye-movement and all subject-level variables, respectively. Indices and beta coefficients are skipped in pseudo-formulas. Model 01 is the baseline. A subject random effect $s_i \sim N(0, \sigma_s^2)$ was present in each GLMM model. All subject variables have been explained in Table 5.3.

ID	N vars		Pseudo-formula	N models	
	em	*sub*		2010	2014
01	0	0	$\text{logit}\,\pi = w.frq$	1	1
02	1	0	$\text{logit}\,\pi = w.frq + em.var + w.frq \times em.var$	7	7
03	1	1	$\text{logit}\,\pi = w.frq + em.var + w.frq \times em.var + id.wm.fl$		7
04	1	1	$\text{logit}\,\pi = w.frq + em.var + w.frq \times em.var + read.spd$		7
05	1	1	$\text{logit}\,\pi = w.frq + em.var + w.frq \times em.var + ctx.fat$		7
06	1	1	$\text{logit}\,\pi = w.frq + em.var + w.frq \times em.var + ctx.pre$		7
07	1	1	$\text{logit}\,\pi = w.frq + em.var + w.frq \times em.var + ctx.crv$		7
08	1	1	$\text{logit}\,\pi = w.frq + em.var + w.frq \times em.var + ctx.crv.3s$		7
09	1	1	$\text{logit}\,\pi = w.frq + em.var + w.frq \times em.var + ctx.totd$		7
10	1	1	$\text{logit}\,\pi = w.frq + em.var + w.frq \times em.var + ctx.totd.3s$		7
11	1	8	$\text{logit}\,\pi = w.frq + em.var + w.frq \times em.var + (all.sub.vars)$		7
12	7	0	$\text{logit}\,\pi = w.frq + (all.em.vars)$	1	1
13	7	8	$\text{logit}\,\pi = w.frq + (all.em.vars) + (all.sub.vars)$		1
TOTAL				9	73

5.2 Results: Single Model Performance

Figures 5.3–5.6 show examples of classification performance plots I used to evaluate performance of my models. These four figures show four combinations of the model family (GLM and GLMM) and the data to which they were fitted (2010 and 2014). I choose to show all four because I think it is important for the reader to get a good idea about how these four combinations looked like before I discuss how I aggregated over them. For brevity I do not report these sort of individual model performance results for all the models because most of them are quite similar; all similarities and differences are the focus of the next subsection.

Each of the Figures 5.3–5.6 contains 20 subplots organized into five columns and four rows. The columns correspond to time windows of increasing size (2–30 seconds) while rows correspond to the four types of model evaluation schemes. As I have discussed in Section 5.1.5, the third and fourth rows are most important. The abscissa of each subplot shows the minimum word length to be considered (measured in character spaces) and the ordinate shows the area under the ROC curve (ROC AUC; higher values indicate better classification performance). The value of 0.5 is marked on the y-axis and indicates a due-to-chance binary classification performance (e.g., one which would result from using a fair coin to perform the classification). There are three series within each subplot, each corresponding to one classification task (see Section 5.1.3).

Several patterns can be seen in Figures 5.3–5.6. First, the GLM model (Figures 5.3 and 5.4) performs essentially at a chance level for both the 2010 and 2014 data and is therefore not worth further discussion. By contrast, the GLMM model (Figures 5.5 and 5.6) performs clearly at an above-chance level when fitted to both the 2010 and 2014 data. One difference between the 2010 and the 2014 data is that distinguishing between normal reading and probe-caught mindless reading (as indicated by the green N-P series) is easier in the 2014 data. At the same time however, the performance on the other two binary classification tasks seems worse and less stable across the different time window sizes and minimum word lengths in the 2014 data as compared to the 2010 data. Finally, while the performance according to the bootstrapped cross-validation (i.e., the third row) seems promising, it is not promising at all according to the subject-stratified bootstrapped cross-validation (i.e., the fourth row). This makes this particular model likely to perform abysmally bad when used to detect mind wandering in readers which eye-movements it has not seen before (i.e., has not been trained on). Similar patterns were present in other individual classification

Table 5.3: Eye-movement and subject-level variables used in classification models investigated.

Variable	Description
$id.wm.fl$	Working memory capacity
$read.spd$	Reading speed
$ctx.fat$	Fatigue
$ctx.pre$	Preoccupation
$ctx.crv$	Craving
$ctx.crv.3s$	Craving (3-state)
$ctx.totd$	Time of the day
$ctx.totd.3s$	Time of the day (3-state)
$em.ffd$	First-fixation duration
$em.gd$	Gaze duration
$em.nb$	Number of blinks
$em.nfpf$	Number of first-pass fixations
$em.nfpr$	Number of first-pass regressions
$em.pd$	Pupil diameter
$em.sfd$	Single fixation duration

performance plots.

5.3 Results: All Models Performance

Comparing the performance of individual models using plots like the ones I have shown in the previous subsection would be challenging. That is why to get a more holistic view of the outcome of my experiment I chose to represent each of those plots as a vector of fewer numbers which I could then easily plot side-by-side. To that end I compressed the individual performance plots by averaging the ROC AUC over the minimum word length and then time window size using the following formulas

$$\text{ROC.AUC}_{ijk} = \Big[\sum_t \Big(\sum_w \text{ROC.AUC}_{ijk.wt}\Big)\frac{1}{|w|}\Big]\frac{1}{|t|},$$
$$\text{IQR}_{ijk} = \Big[\sum_t \Big(\sum_w \text{Q}_{ijk.wt.3} - \text{Q}_{ijk.wt.1}\Big]\frac{1}{|w|}\Big)\frac{1}{|t|},$$

where i iterates through classification task (see Section 5.1.3), j iterates through evaluation scheme (see Section 5.1.5), k iterates through models (see Section 5.1.8), $t \in \{2, 5, 10, 20, 30\}$ is the time window size, $w \in \{8, 10, 12\}$ is the word length, IQR denotes interquantile range, and Q denotes quantile. Thus, I obtained nine numbers per row of an individual classification performance plot (or 36 numbers per plot in total). Note that because, as I have shown in Section 4.11, differences in values of eye-movement variables are largest for long words, I only used words eight character spaces or longer when aggregating.

5.3.1 GLM

Figures 5.7 and 5.8 show the performance of all the GLM family classifiers with data aggregated over all five time windows. All models fitted to the 2014 data perform essentially at the chance level and thus cannot reliably discern between normal and mindless reading. The same is true for all but one model fitted to the 2010 data. That one exception is model *02-em.pupil-0* which includes pupil diameter. This could be related to the observation that pupil size is lower during probe-caught mindless reading relative to normal reading (see Section 4.11; note though that this result is for the 2014 data). Additionally, within-subject variability in the 2010 data is estimated well thanks to the long reading times. Therefore, this could further explain why pupil diameter still helped in estimating the readers attentional state. Yet another plausible explanation has to do with how pupil diameter is recorded by the EyeLink 1000 eye tracker. The eye tracker does not register absolute sizes (e.g., in millimeters); instead, each subject's pupil size data has to be compared with their own reference size recorded by the experimenter. These reference sizes were not collected for the 2010 data (nor for the 2014 data, by the way) but the small number of subjects could have offset that fact because of small between-subject variation. The GLMM family of models, to which I turn next, allows to sidestep the need for reference size altogether by estimating each subject's individual contribution.

5.3.2 GLMM

Figures 5.9–5.11 respectively show the performance of fast (i.e., 2 and 5 seconds), medium (i.e., 10 seconds), and slow (i.e., 20 and 30 seconds) GLMM classifiers used on the 2010 data. The story is similar for all three reaction speeds and therefore I describe them together. Looking at the third row of all the figures (which shows results of 10-fold cross-validation), the baseline model (i.e., model *01* shown in the first column) performs well as indicated by the ROC AUC at roughly 0.6 level. Surprisingly, addition of an eye-movement variable to that model does not always translate into performance boost. In fact, some of the models perform on a sub-baseline level. Nevertheless, the differences, regardless of direction, are quite small. As expected, the most structurally complex model (i.e, model *12*) does the worst. Model including pupil diameter is at the other end of the performance spectrum achieving the best classification performance that reaches about 0.8. While differences across the three reaction speeds (i.e., fast, medium, and slow) are small, the fast pupil-diameter model is the best overall. Looking across classification tasks (i.e., N-P, N-S, and N-PS), discriminating between normal reading and probe-caught mindless reading (i.e., N-P) is the easiest of the three. Moving down to row four

(which shows results of 10-fold leave-subject-out cross-validation), the pupil diameter model stands out even more achieving ROC AUC of nearly 0.75 which demonstrates that this classifier is expected to do well on classifying an attentional state of a reader it has not been trained on.

Figures 5.12–5.14 respectively show the performance of fast, medium, and fast GLMM classifiers used on the 2014 data. Looking at the third row of all the figures, just like in the 2010 results, the N-P task is also the easiest here. However, there are bigger differences between the three reaction speeds then was the case for the 2010 results. More specifically, overall, as expected, the fast classifiers perform the worst (but several still achieve good ROC AUC of over 0.75) and the slow ones the best (nearly all have ROC AUC above 0.75). Comparing models within reaction speed, the medium and fast baseline model (i.e., models *01*) do surprisingly well. The fast baseline model does a bit worse than other fast models. Still, as was the case with the 2010 data, addition of variables (be it subject-level or eye-movement) can both deteriorate and improve classification performance. Evidently, no one variable seems to consistently affect performance which suggests that combinations of subject-level and eye-movement variables work together, with some combinations being better than others. The 2014 results can be used to select the best classifier within each reaction speed but I refrain from naming the winners because many models are comparable. Moving on to row four, several classifiers, especially among the medium ones, stand out promising to do well on new eye-movement data. However, the vast majority of models cannot be relied on.

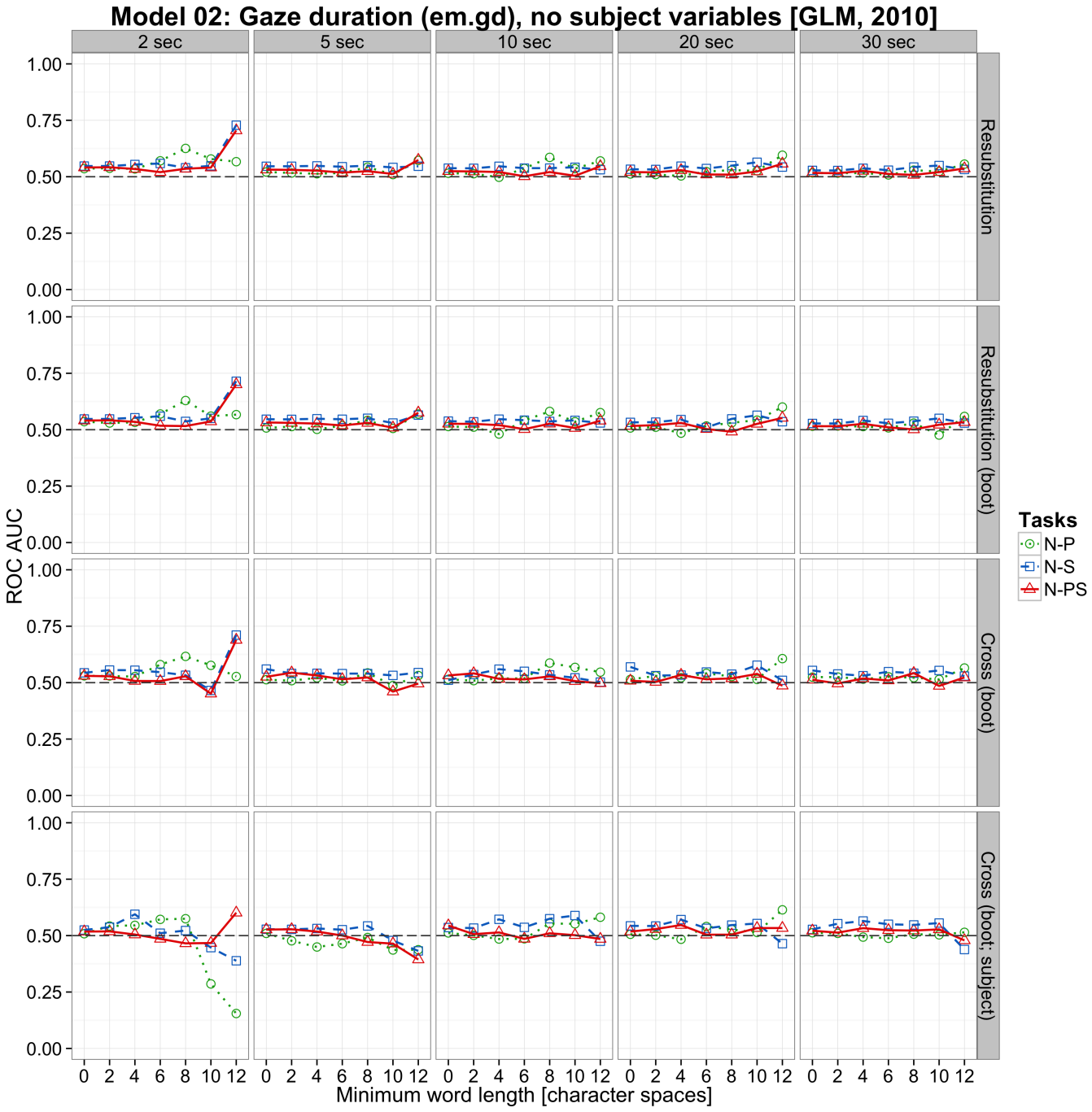


Figure 5.3: Classification performance of the GLM model 02 (Table 5.2) fitted to the 2010 data.

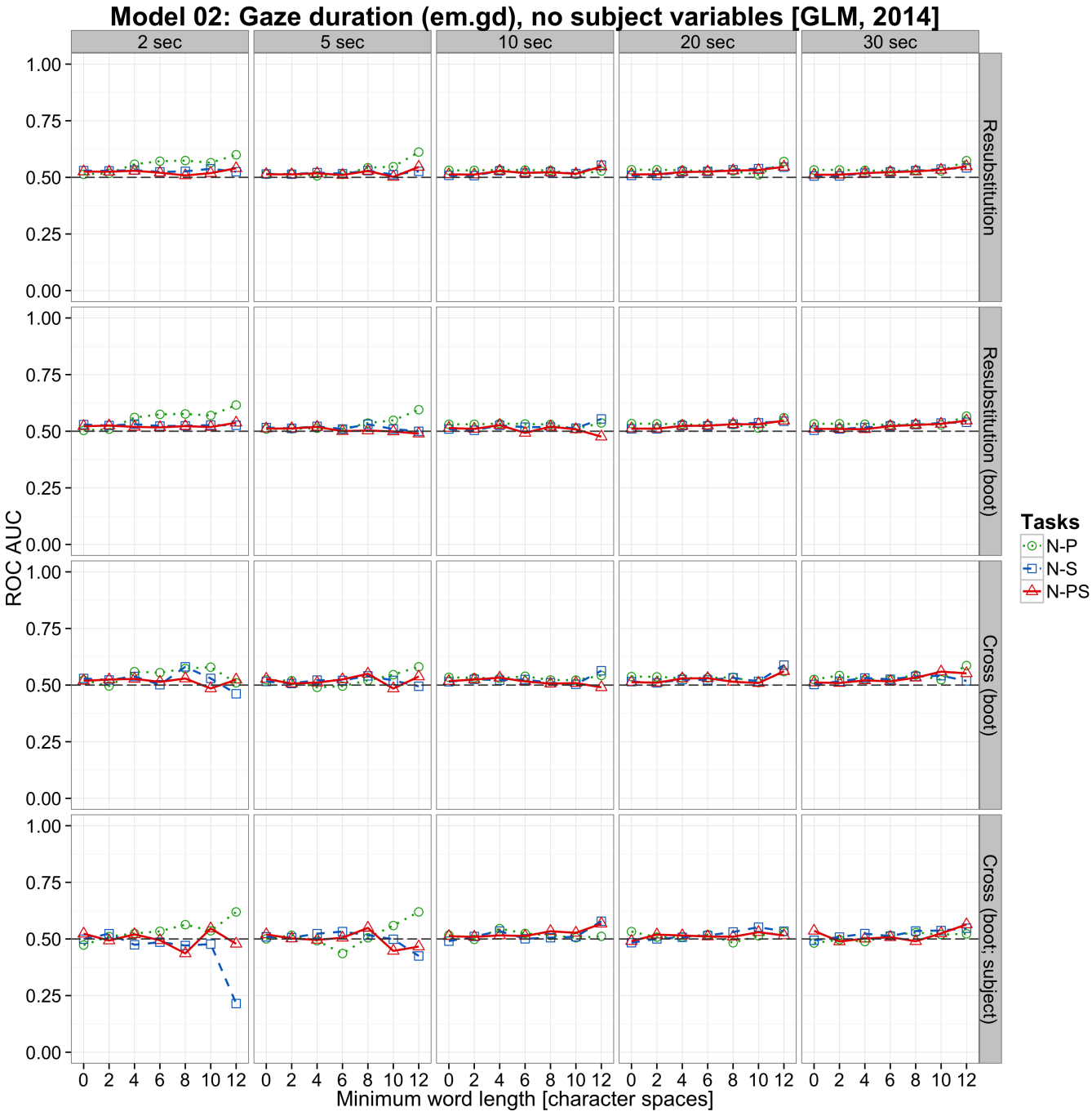


Figure 5.4: Classification performance of the GLM model 02 (Table 5.2) fitted to the 2014 data.

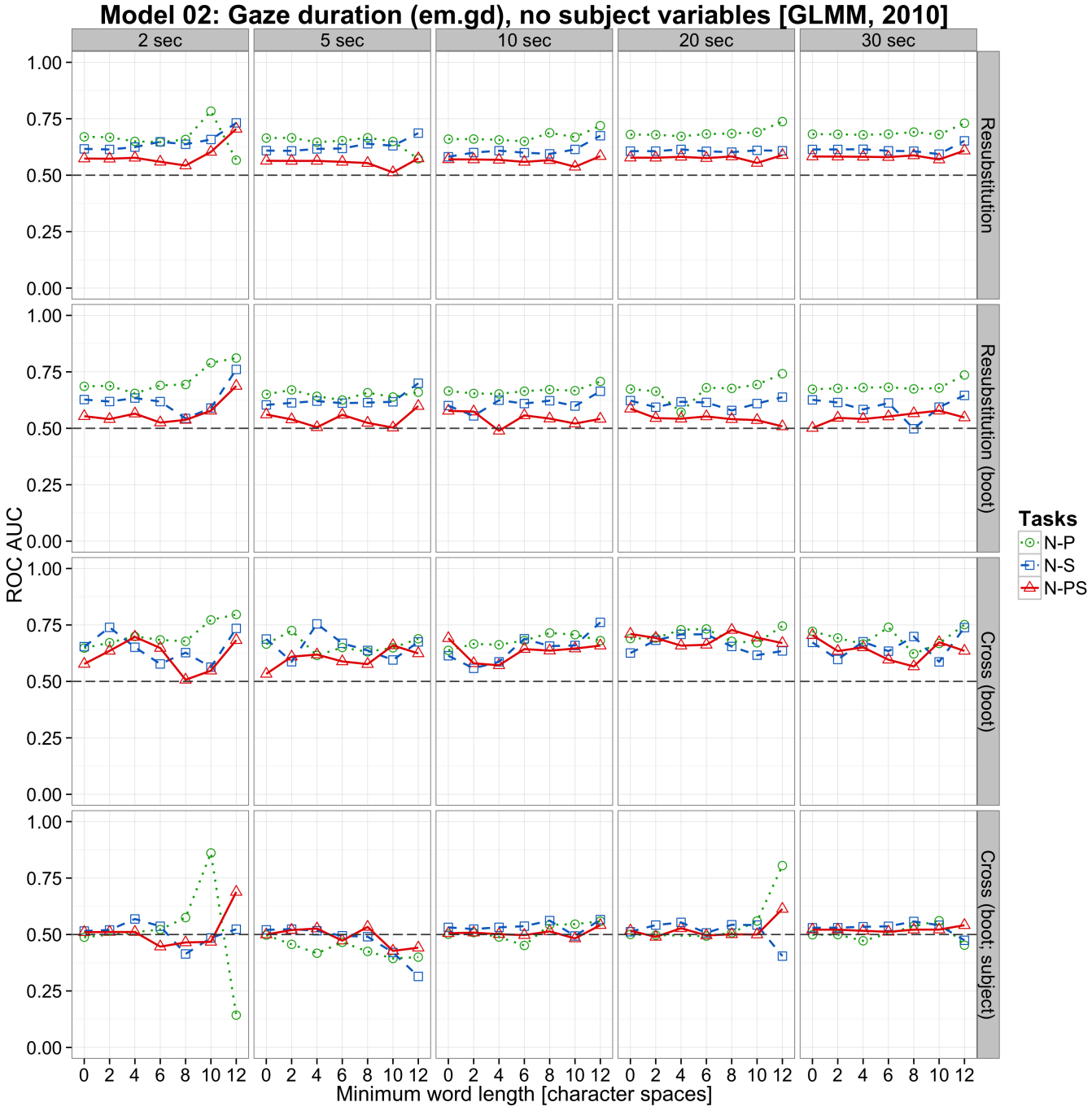


Figure 5.5: Classification performance of the GLMM model 02 (Table 5.2) fitted to the 2010 data.

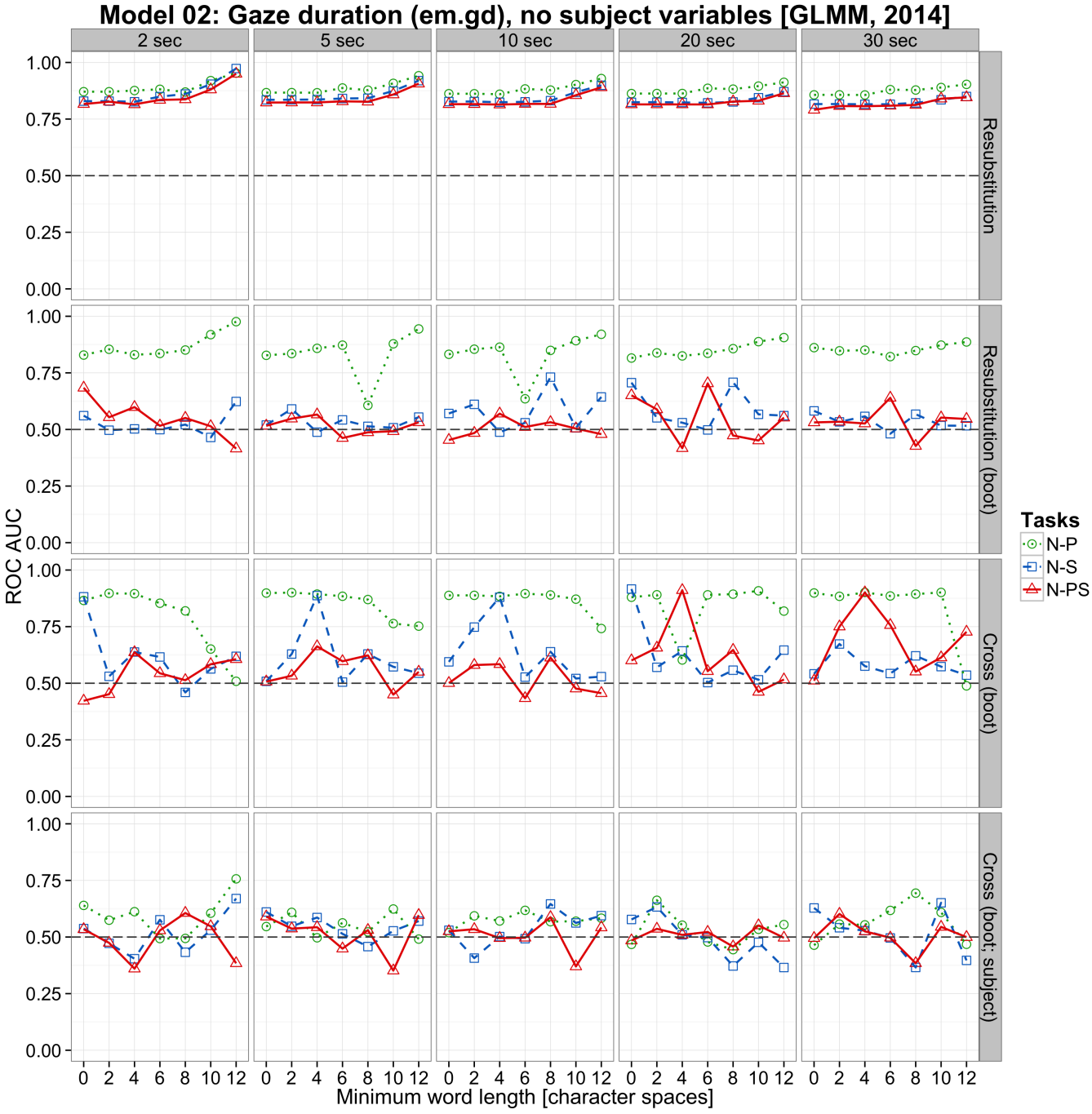


Figure 5.6: Classification performance of the GLMM model 02 (Table 5.2) fitted to the 2014 data.

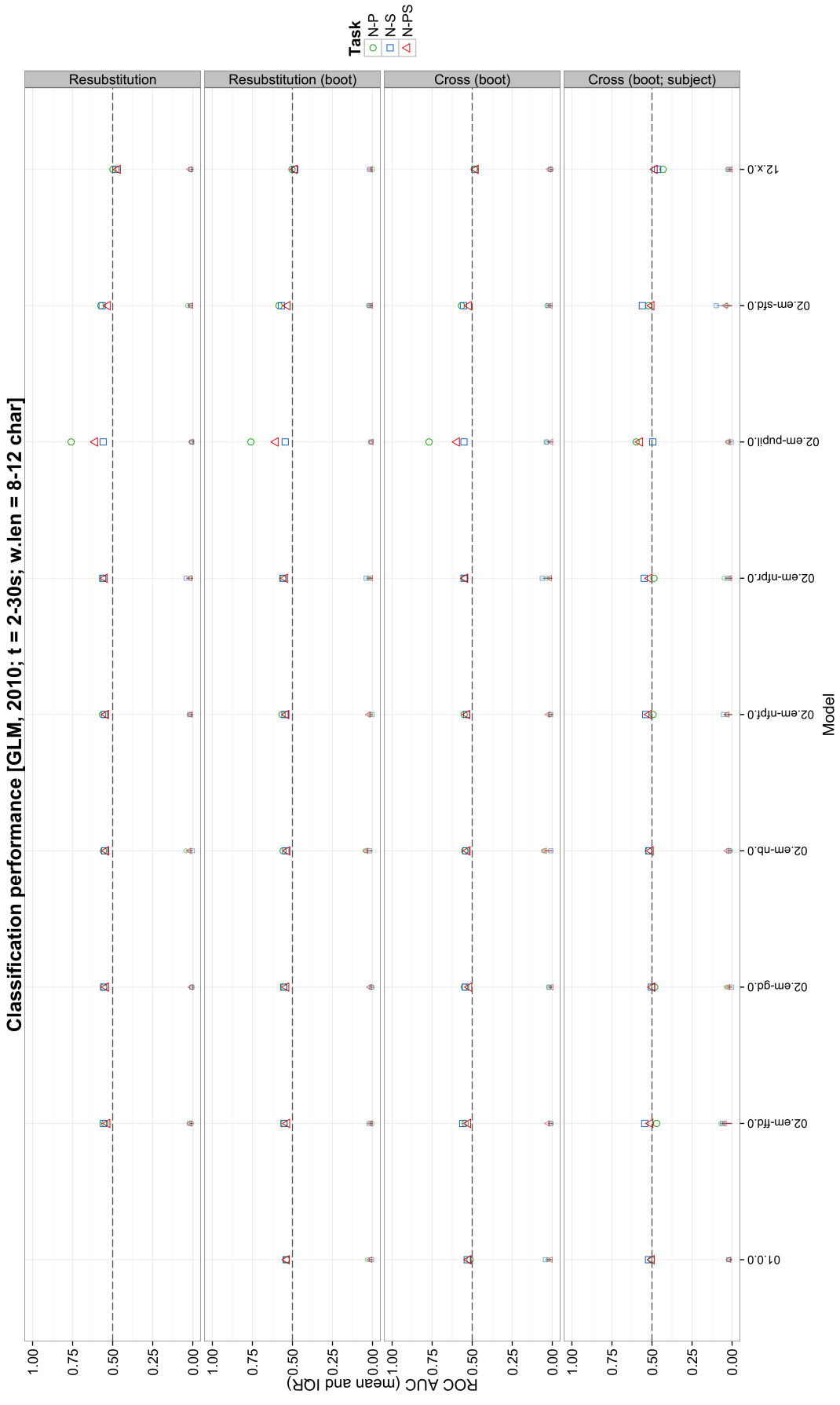


Figure 5.7: Classification performance of GLM models fitted to the 2010 data. Vertical axis shows mean and interquantile range (IQR) of the area under the ROC curve (ROC AUC). Data from all time windows and words longer than eight character spaces.

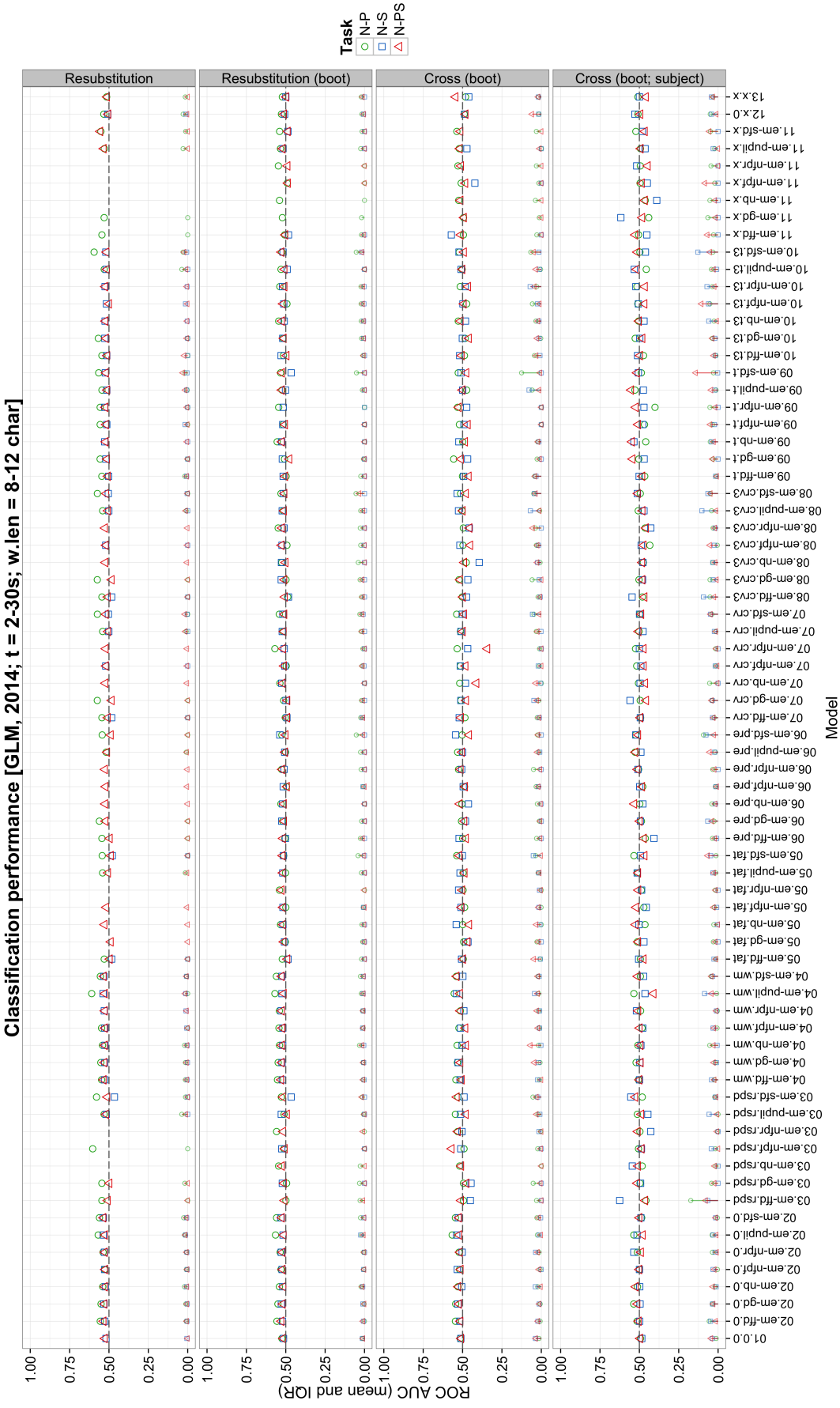


Figure 5.8: Classification performance of GLM models fitted to the 2014 data. Vertical axis shows mean and interquantile range (IQR) of the area under the ROC curve (ROC AUC). Data from all time windows and words longer than eight character spaces.

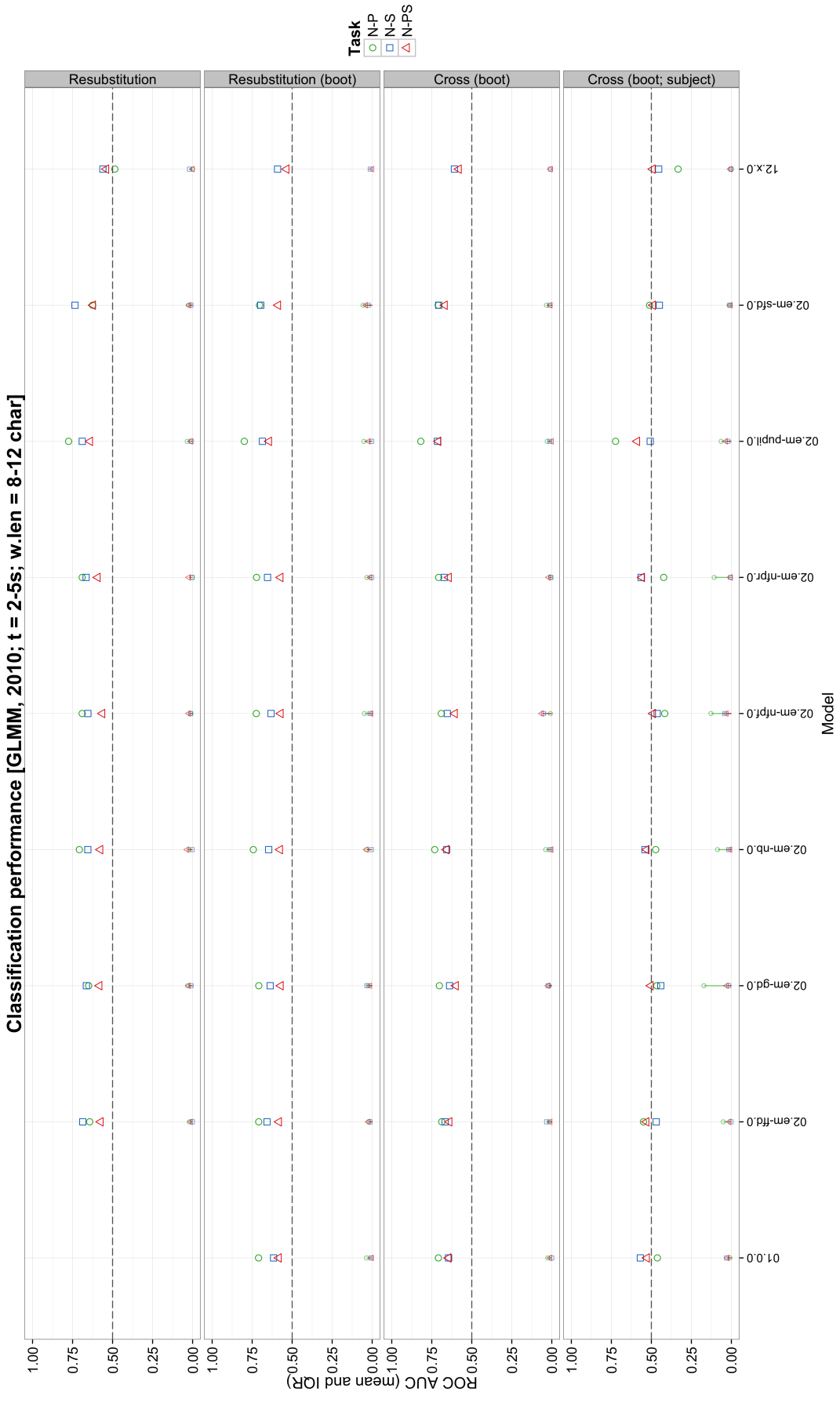


Figure 5.9: Classification performance of GLMM models fitted to the 2010 data. Vertical axis shows mean and interquantile range (IQR) of the area under the ROC curve (ROC AUC). Data from the 2s and 5s time windows (i.e., fast classifiers) and words longer than eight character spaces.

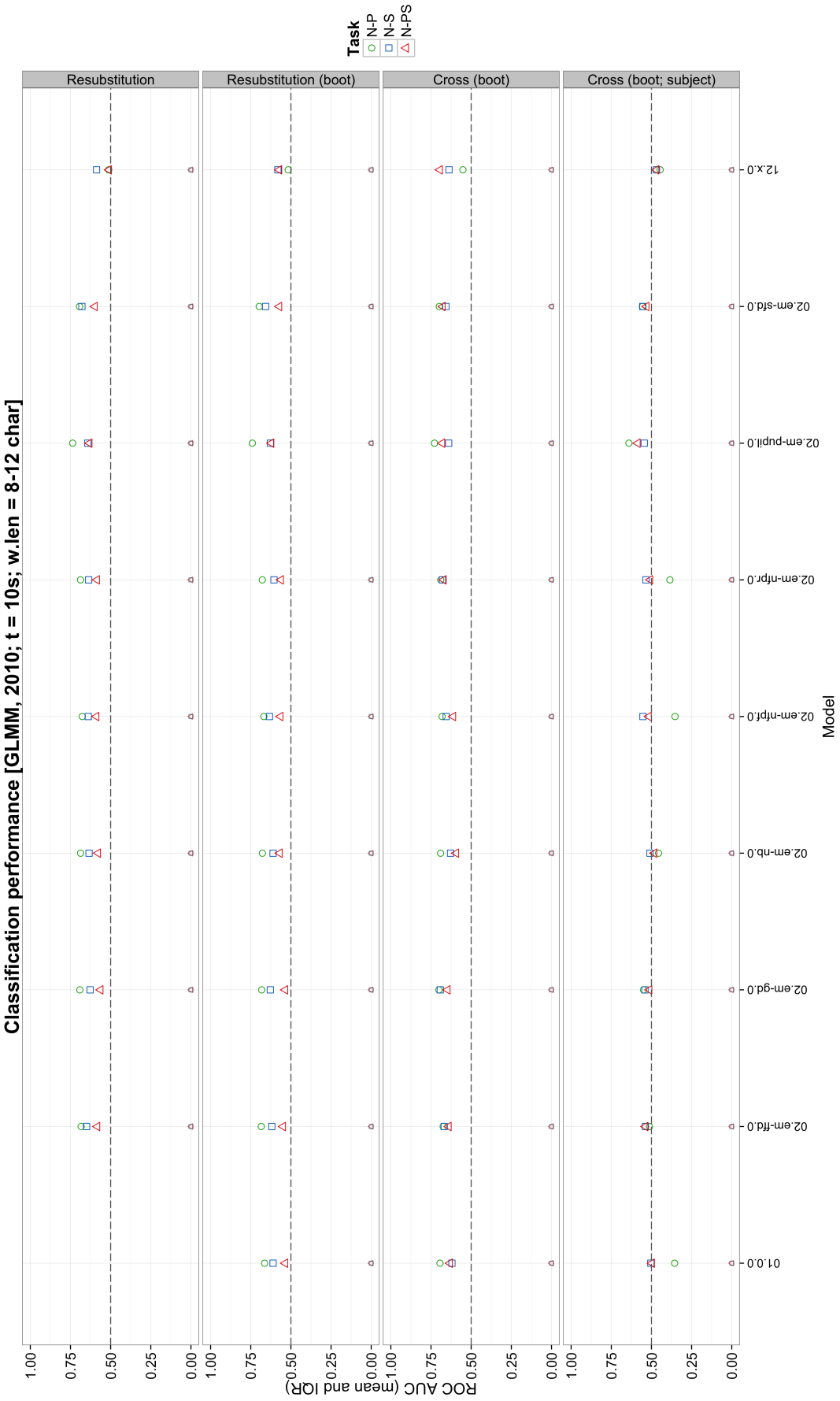


Figure 5.10: Classification performance of GLMM models fitted to the 2010 data. Vertical axis shows mean and interquantile range (IQR) of the area under the ROC curve (ROC AUC). Data from the 10s time window (i.e., medium classifiers) and words longer than eight character spaces.

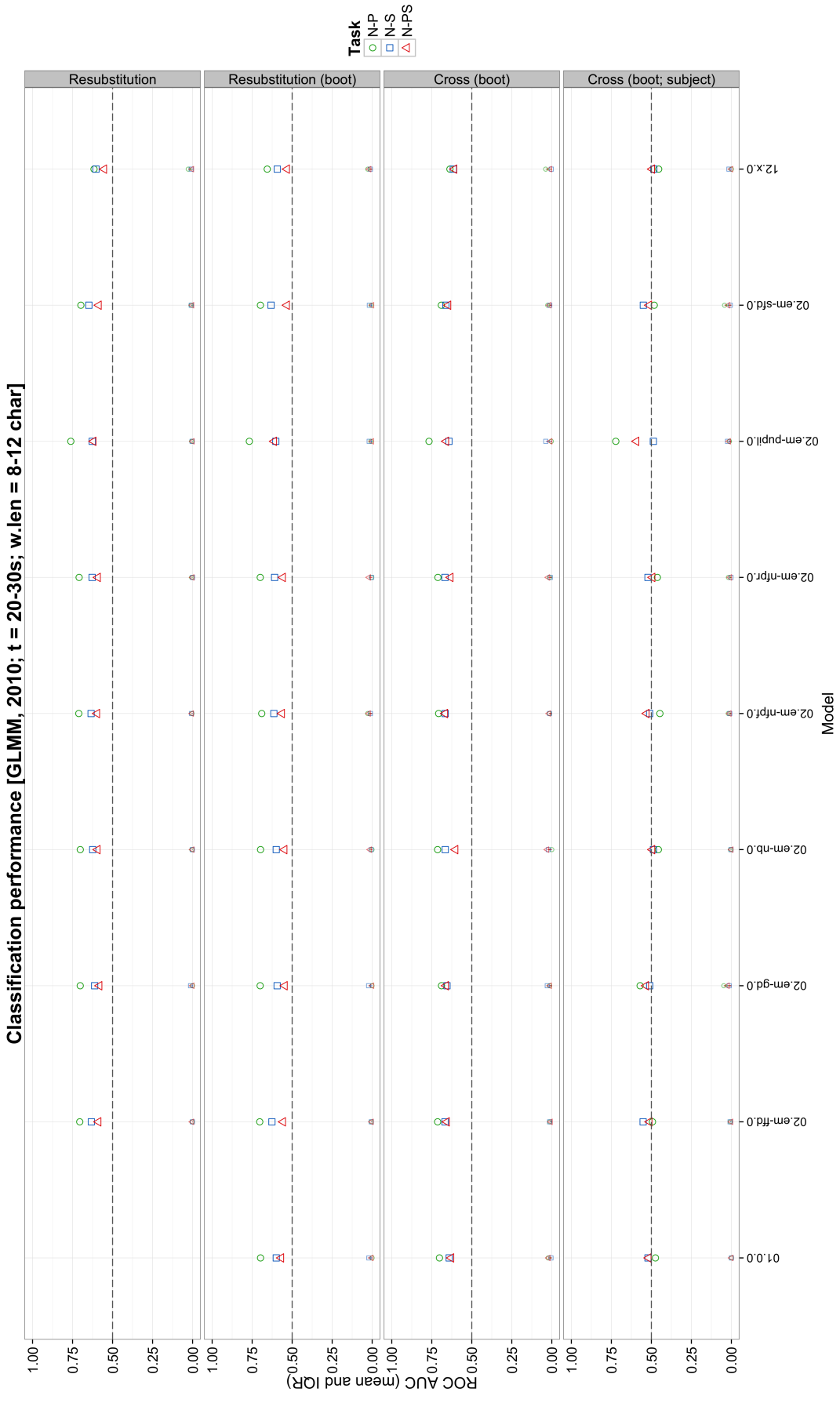


Figure 5.11: Classification performance of GLMM models fitted to the 2010 data. Vertical axis shows mean and interquantile range (IQR) of the area under the ROC curve (ROC AUC). Data from the 20s and 30s time windows (i.e., slow classifiers) and words longer than eight character spaces.

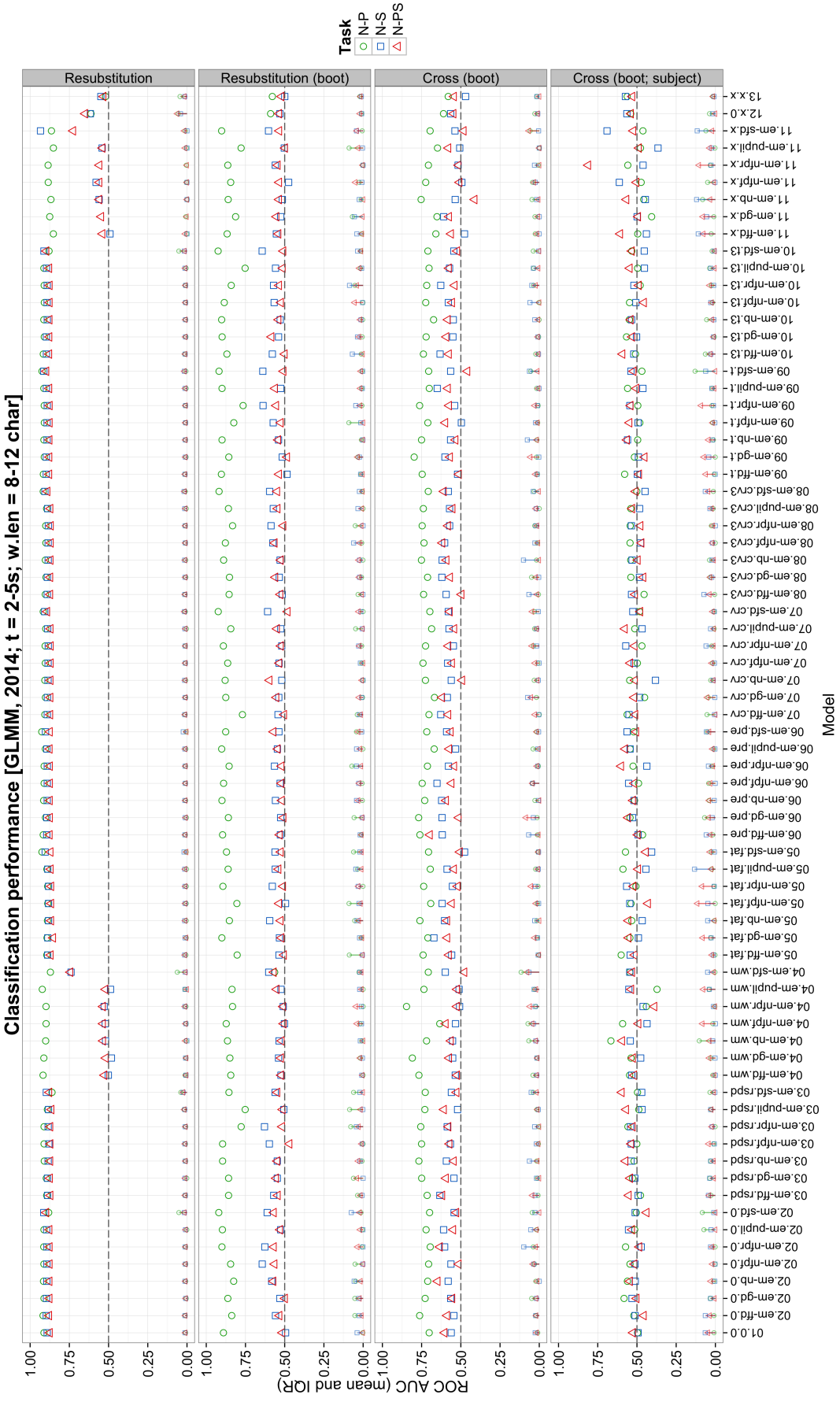


Figure 5.12: Classification performance of GLMM models fitted to the 2014 data. Vertical axis shows mean and interquantile range (IQR) of the area under the ROC curve (ROC AUC). Data from the 2s and 5s time windows (i.e., fast classifiers) and words longer than eight character spaces.

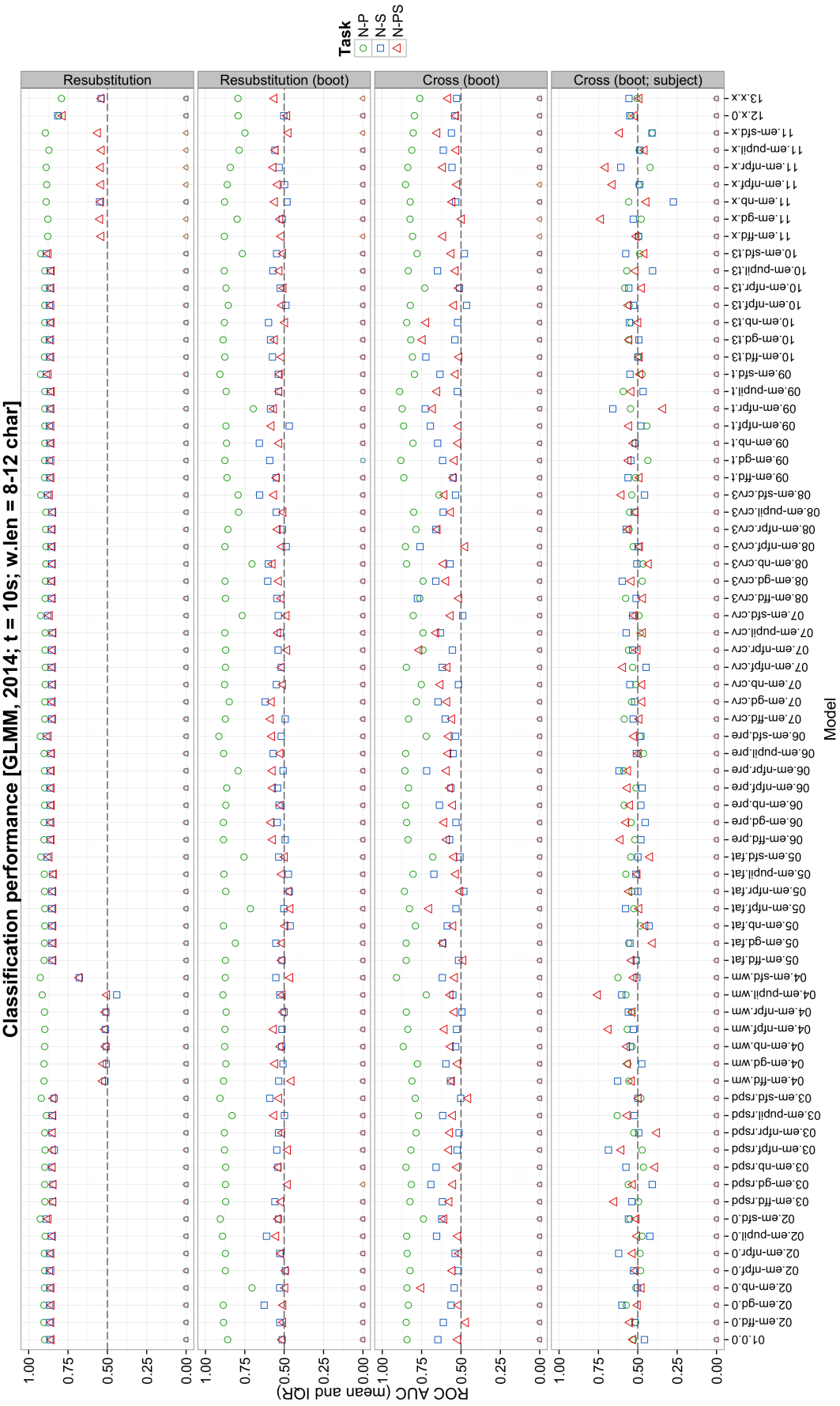


Figure 5.13: Classification performance of GLMM models fitted to the 2014 data. Vertical axis shows mean and interquantile range (IQR) of the area under the ROC curve (ROC AUC). Data from the 10s time window (i.e., medium classifiers) and words longer than eight character spaces.

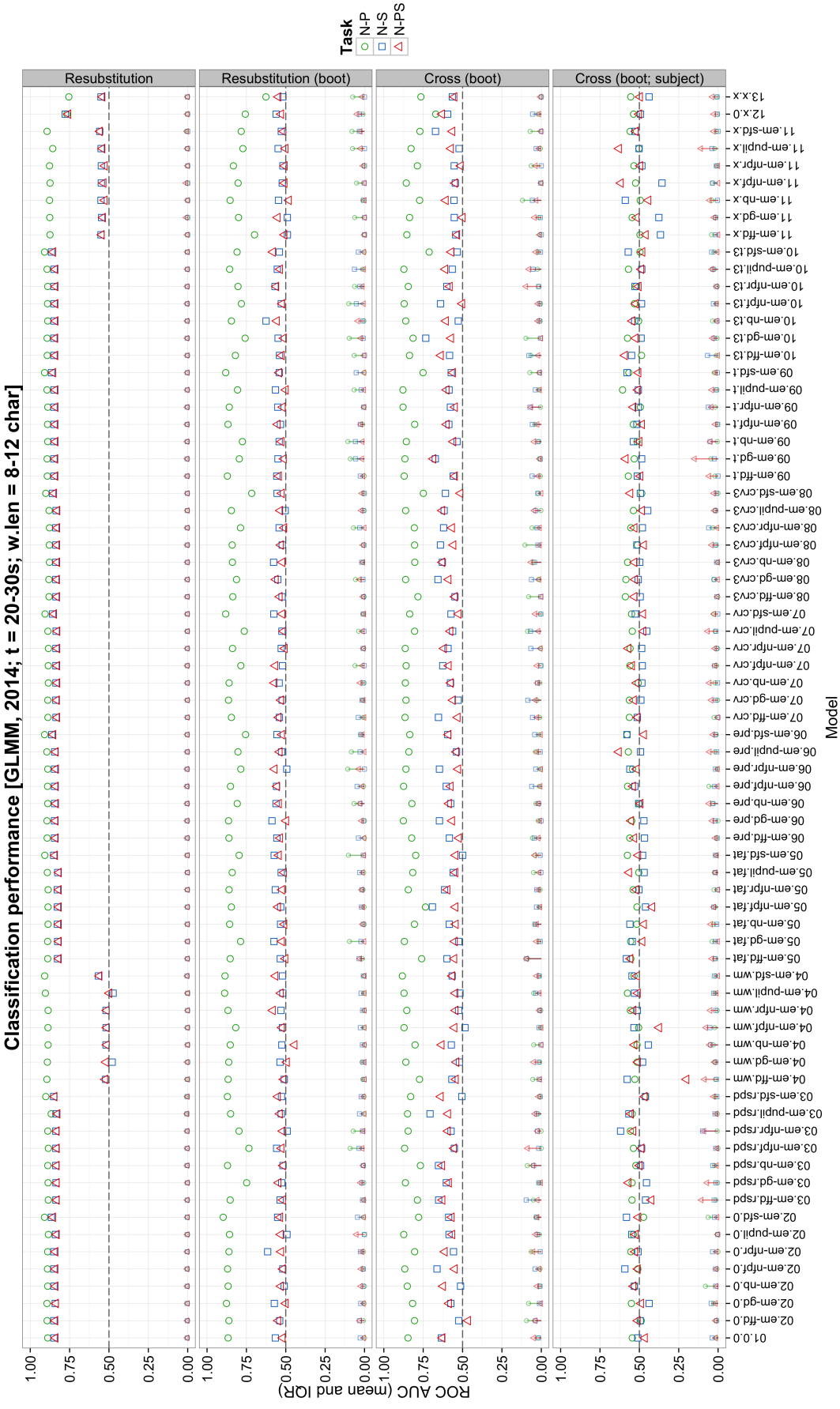


Figure 5.14: Classification performance of GLMM models fitted to the 2014 data. Vertical axis shows mean and interquantile range (IQR) of the area under the ROC curve (ROC AUC). Data from the 20s and 30s time windows (i.e., slow classifiers) and words longer than eight character spaces.

Chapter 6

Conclusions

This document describes my analyses of the data from the largest mindless reading experiment run to date. I performed these analyses with two goals in mind: (1) Study and (2) detection of mindless reading. In the study part of this work (Chapter 4), I have discovered that the incidence of mindless reading was higher for readers who found the text less interesting as well as for those who were more fatigued. Furthermore, I have found that readers who read faster and were more preoccupied were more likely to self-report mindless reading. Moreover, after scrutinizing reading speed differences in mindful versus mindless reading I have found evidence consistent with the hypothesis that effects of lexical variables are attenuated during mindless reading. That hypothesis was further reinforced by my investigation of the relationship between mindless reading, text comprehension, and working memory, the results of which suggest that text comprehension suffers as a result of mind-wandering.

In my analyses involving eye-movements, I have found a higher incidence of off-screen fixations during mindless reading. Furthermore, I have found that extremely short fixations (i.e., those shorter than 80ms) were less probable in probe-caught mindless reading but more probably during self-caught mindless reading. Additionally, extremely long fixations (i.e., those longer than 1000ms) were more probable during probe-caught mindless reading and even more so during self-caught mindless reading.

My analysis of the influence of the previous and next words' variables on eye-movements on the current word revealed an anticipated strong immediacy effects as well as faint and uncontroversial lag and successor effects. Moreover, I have found evidence for shallower text processing during mindless reading as indicated by stronger immediacy perceptual effect and weaker lexical effect. Importantly, I have not detected the controversial frequency successor effect. However, I have observed a weird "anti-spillover" effect the presence of which may be indigenous to circumstances involving natural reading of extended corpora of text.

In the detection part of this dissertation (Chapter 5), I have demonstrated that word-level and eye-movement variables from a record of an ecologically valid task of reading extended passages of text coming from two different experiments (which I refer to as 2010 and 2014) can be used to disentangle instances of mindless reading from the intervals of mindful reading. While only proofs-of-concept, the classification models I evaluated do well in the three binary classification tasks I chose for them. Discriminating between normal reading and probe-caught mindless reading is the least challenging of these tasks. This implies that probe-caught mindless reading differs from normal reading more so than self-caught mindless reading, which further highlights that probe-caught mindless reading likely is the purest or deepest form of mind-wandering a reader can lapse into. Naturally, this also means that if the models I have discussed were to replace the current state-of-the-art method of discovering mindless reading their utility will be the highest in experiments which aim at catching readers mindlessly reading without meta-awareness.

In several places throughout this document I have alluded to potential differences between normal reading and self-caught mindless reading. Many of these differences are manifested right before a self-report of mindless reading. While detecting the moments of regaining meta-awareness has not been my goal, it is possible that these differences can be used to discern normal reading and self-caught mindless reading more effectively than I have been able to do with the approach I have taken.

There is limited evidence that the classification models I have investigated can handle eye-movement data from subjects they had not seen before. As a result, in order to use these models, one would need to obtain a sample of eye-movement data from a subject prior to attempting to identify mindless reading in that subject. That is, a kind of a burn-in reading session would be necessary. However, because these results imply that there are differences between readers which these models do not capture, more research is necessary to properly address this limitation.

The good performance of the baseline model in both the 2010 and 2014 data is not something I expected. It appears that knowing which words a reader fixates is enough to

warrant good classification performance. At the same time, the results suggest that adding subject-level or an eye-movement variables can both harm and improve that performance, but there does not seem to be any clear indication on which variables do what. Furthermore, the results for the 2014 data suggest that subject-level and eye-movement variables work together and that some of these combinations are better than others.

While the 2010 results suggest that accounting for pupil size improves classification performance, it should be noted that the 2010 data is based on only four subjects. As a result, generalizations to the universe of readers should be made with caution.

Finally, the distinct difference between the performance of GLM and GLMM models I have found show that neglecting to account for correlation between observations coming from the same reader may result in abysmal classification performance.

Extensions of this work include deploying the classification models I have tested in reading experiments to determine how well they perform "in the field." Of course, that implies the development of an interface between these models and an eye tracker; that interface would feed the models with a real-time stream of eye movements. Furthermore, more structurally complex models (i.e., those utilizing more variables or more relationships between variables) could achieve better classification performance.

Bibliography

Altarriba, J., Kambe, G., Pollatsek, A., & Rayner, K. (2001) Semantic codes are not used in integrating information across eye fixations in reading: Evidence from fluent Spanish-English bilinguals. *Perception & Psychophysics*, 63, 875–890.

Antrobus, J.S. (1968) Information theory and stimulus-independent thought. *British Journal of Psychology*, 59, 423–430.

Antrobus, J. S., Singer, J. L., Goldstein, S., & Fortgang, M. (1970) Mindwandering and cognitive structure. *Transactions of the New York Acedemy of Science, Series II*, 32, 242–252.

Antrobus, J. S., Singer, J. L., & Greenberg, S. (1966) Studies in the stream of consciousness: Experimental enhancement and suppression of spontaneous cognitive process. *Perceptual and Motor Skills*, 22, 399–417.

Balota, D.A., Pollatsek, A., & Rayner, K. (1985) The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, 17, 364–390.

Barrouillet, P. (1996) Transitive inferences from set-inclusion relations and working memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22, 1408–1422.

Beatty, J. (1982) Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276–92.

Beatty, J. & Lucero-Wagoner, B. (2000) The pupillary system. In Cacioppo, J.T., Tassinary, L.G., & Berntson, G.G. (Eds.), *Handbook of Psychophysiology*. Cambridge University Press.

Benton, S.L., Kraft, R.G., Glover, J.A., & Plake, B.S. (1984) Cognitive capacity differences among writers. *Journal of Educational Psychology*, 76, 820–834.

Binder, K.S., Pollatsek, A., & Rayner, K. (1999) Extraction of information to the left of the fixated word in reading. *Journal of Experimental Psychology: Human Perception & Performance*, 25, 1162–1172.

Bleckley, M.K., Durso, F.T., Crutchfield, J.M., Engle, R.W., & Khanna, M.M. (2004) Individual differences in working memory capacity predict visual attention allocation. *Psychonomic Bulletin & Review*, 10, 884–889.

Brown, A.L. (1980) Metacognitive development and reading. In Spiro, R.J., Bruce, B.C., & Brewer, W.F. (Eds.), *Theoretical issues in reading comprehension* (pp. 453–481). Hlllsdale, NJ Lawrence Erlbaum Associates.

Brysbaert, M. & New, B. (2009) Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41 (4), 977-990.

Cabeza, R., Prince, S.E., Daselaar, S.M., Greenberg, D.L., Budde, M., Dolcos, F., LaBar, K.S., & Rubin, D.C. (2004) Brain activity during episodic retrieval of autobiographical and laboratory events: an fMRI study using a novel photo paradigm. *Journal of Cognitive Neuroscience*, 16(9), 1583–1394.

Cameron, A. C. & Trivedi, P. K. (1986) Econometric models based on count data: Comparisons and applications of some estimators. *Journal of Applied Econometrics*, 1, 29-53.

Carpenter, P.A. & Daneman, M. (1981) Lexical retrieval and error recovery in reading: A model based on eye fixations. *Journal of Verbal Learning & Verbal Behavior*, 28, 138–160.

Chaffin, R., Morris, R.K., & Seely, R.E. (2001) Learning new word meanings from context: A study of eye movements. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 27, 225–235.

Clarkson-Smith, L. & Hartley, A.A. (1990) The game of bridge as an exercise in working memory and reasoning. *Journal of Gerontology*, 45, 233–238.

Clifton, C., Staub, A., & Rayner, K. (2007) Eye movements in reading words and sentences. *Eye movements: A window on mind and brain* (pp. 341–371), New York: Elsevier.

Coltheart, M. (1981) The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, 33A, 497–505.

Consul, P.C. (1989) *Generalized poisson distribution: Properties and applications*. Marcel Dekker, New York, New York, USA.

Conway, A.R.A., Kane, M.J., & Engle, R.W. (2003) Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, 7, 547–552.

Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O., & Engle, R.W. (2005) Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769–786.

Cunningham, S., Scerbo, W., & Freeman, F. (2000) The electrocortical correlates of daydreaming during vigilance tasks. *The Journal of Mental Imagery*, 24, 61–72.

D'Mello, S., Cobian, J., & Hunter, M. (2013) Automatic gaze-based detection of mind wandering during reading. *Proceedings of the 6th International Conference on Educational Data Mining (EDM)*, 364–365.

Daily, L.Z., Lovett, M.C., & Reder, L.M. (2001) Modeling individual differences in working memory performance: A source activation account. *Cognitive Science: A Multidisciplinary Journal*. 25(3), 315–353.

Daneman, M. & Carpenter, P.A. (1980) Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, 19, 450–466.

Daneman, M. & Carpenter, P.A. (1983) Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 9, 561–584.

Daneman, M. & Green, I. (1986) Individual differences in comprehending and producing words in context. *Journal of Memory & Language*, 25, 1–18.

Daneman, M. & Merikle, P.M. (1996) Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, 3, 422–433.

Den Buurman, R., Boersma, T., & Gerrissen, J.F. (1981) Eye movements and the perceptual span in reading. *Reading Research Quarterly*, 16, 227–235.

Dougherty, M.R.P. & Hunter, J. (2003) Probability judgment and subadditivity: The role of working memory capacity and constraining retrieval. *Memory & Cognition*, 31, 968–982.

Duffy, S.A., Morris, R.K., & Rayner, K. (1988) Lexical ambiguity and fixation times in reading. *Journal of Memory & Language*, 27, 429–446.

Efron B. (1983) Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association*, 78, 316-331.

Ehrlich, S.E. & Rayner, K. (1981) Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning & Verbal Behavior*, 20, 641–655.

Engbert, R., Longtin, A., & Kliegl, R. (2002) A dynamic model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, 42(5), 621–36.

Engbert, R., Nuthmann, A., Richter, E.M., & Kliegl, R. (2005) SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, 112(4), 770–813.

Engle, R.W., Carullo, J.J., Collins, & K.W. (1991) Individual differences in working memory for comprehension and following directions. *Journal of Educational Research*, 84, 253–262.

Engle, R.W., Tuholski, S.W., Laughlin, J.E., & Conway, A.R.A. (1999) Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, 128, 309–331.

Findlay, J.M. & Gilchrist, I.D. (2003) Active vision: The psychology of looking and seeing. Oxford University Press.

Folk, J.R. (1999) Phonological codes are used to access the lexicon during silent reading. Journal of Experimental Psychology: Learning, Memory, & Cognition, 25, 892–906.

Giambra, L. M., Rosenberg, E. H., Kasper, S., Yee, W., & Sack, D. A. (1988) A circadian rhythm in the frequency of spontaneous task-unrelated images and thoughts. *Imagination, Cognition, and Personality*, 8, 307–312.

Giambra, L. M. (1989) Task-unrelated thought frequency as a function of age: A laboratory study. *Psychology and Aging*, 4, 136-143.

Giambra, L.M. (1995) A laboratory based method for investigating influences on switching attention to task unrelated imagery and thought. *Consciousness & Cognition*, 4, 1–21.

Geisser, S. (1993) Predictive Inference. New York: Chapman and Hall.

Glenberg, A.M., Wilkinson, A., & Epstein, W. (1982) The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, 10, 597–602.

Grodsky, A., & Giambra, L. (1989) Task unrelated images and thoughts whilst reading. In J. Shorr, P. Robin, J.A. Connek, & M. Wolpin (Eds.), Imagery: Current perspectives. New York: Plenum Press.

Hacker, D.J. (1998) Self-regulated comprehension during normal reading. In Hacker, D.J., Dunlosky, J., & Graesser, A.C. (Eds.), *Metacognition in educational theory and practice* (pp. 165–191). Mahwah, NJ: Erlbaum.

Heitz, R.P. & Engle, R.W. (2007) Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology: General*, 136, 217-240.

Henderson, J.M. & Ferreira, F. (1990) Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 16, 417–29.

Henderson, J.M. & Ferreira, E. (1993) Eye movement control during reading: Fixation measures reflect foveal but not parafoveal processing difficulty. *Canadian Journal of Experimental Psychology*, 47, 201–221.

Huey, E.B. (1908) The psychology and pedagogy of reading. Macmillan.

Hurlburt, R.T. (1993) Sampling inner experience in disturbed affect. New York: Plenum Press.

Hyönä, J. & Pollatsek, A. (1998) Reading Finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception & Performance*, 24, 1612–1627.

Inhoff, A.W. & Rayner, K. (1986) Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, 40, 431–439.

Inhoff, A.W. & Radach, R. (1998) Definition and computation of oculomotor measures in the study of cognitive processes. In Underwood, G. (Ed.), Eye guidance in reading and scene perception (pp. 29–53). Amsterdam: Elsevier.

Japkowicz N. & Stephen, S. (2002) The class imbalance problem: A systematic study. *Journal of Intelligent Data Analysis*, 6(5), 429–449.

Juhasz, B.J. (2005) Age-of-acquisition effects in word and picture processing. *Psychological Bulletin*, 131, 684–712.

Juhasz, B.J. & Rayner, K. (2003) Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 29, 1312–1318.

Juhasz, B.J. & Rayner, K. (2006) The role of age-of-acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, 13(7-8), 846-863.

Juhasz, B., Starr, M.S., Inhoff, A.W., & Placke, L. (2003) The effects of morphology on the processing of compound words: Evidence from naming, lexical decision, and eye fixations. *British Journal of Psychology*, 94, 233–244.

Just, M.A. & Carpenter, P.A. (1980) A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354.

Just, M.A. & Carpenter, P.A. (1987) The psychology of reading and language comprehension. Allyn and Bacon.

Just, M.A. & Carpenter, P.A. (1992) A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.

Just, M.A., Carpenter, P.A., Keller, T.A., Eddy, W.F., & Thulborn, K.R. (1996). Brain activation modulated by sentence comprehension. *Science, 274*, 114–116.

Just, M.A., Carpenter, P.A., & Woolley J.D. (1982) Paradigms and Processes in Reading Comprehension. *Journal of Experimental Psychology: General*, 3(2), 228–238.

Kane, M.J. & Engle, R.W. (2000) Working memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 26, 333–358.

Kane, M.J., Bleckley, M.K., Conway, A.R.A., & Engle, R.W. (2001) A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, 130, 169–183.

Kane, M.J. & Engle, R.W. (2003) Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132, 47–70.

Kane, M.J., Brown, L.H., McVay, J.C., Silvia, P.J., Myin-Germeys, I., & Kwapil, T.R. (2007) For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science*, 18, 614–621.

Kennison, S.M., Clifton, C. (1995) Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 68–81.

Kiewra, K.A. & Benton, S.L. (1988) The relationship between information processing ability and notetaking. *Contemporary Educational Psychology*, 13, 33–44.

Kincaid, J.P., Fishburne, R.P., Rogers, R.L., & Chissom, B.S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease formula) for Navy enlisted personnel. Research Branch Report 8- 75. Chief of Naval Technical Training: Naval Air Station Memphis.

King, J. & Just, M.A. (1991) Individual differences in syntactic processing: The role of working memory. *Journal of Memory & Language*, 30, 580–602.

Kintsch, W., Healy, A.F., Hegarty, M., Pennington, B.F., & Salthouse, T.A. (1999) Eight questions and some general issues. In Miyake, A. & Shah, P. (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge: Cambridge University Press.

Kliegl, R. (2007) Toward a perceptual-span theory of distributed processing in reading: A reply to Rayner, Pollastsek, Drieghe, Slattery, and Reichle (2007). *Journal of Experimental Psychology: General, 136*, 530–537.

Kliegl, R. & Engbert, R. (2003) SWIFT explorations. In *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 391–411). Elsevier.

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*, 135(1), 12–35.

Klinger, E. (1971) Structure and functions of fantasy. Wiley, New York.

Klinger, E.C. (1999) Thought flow: Properties and mechanisms underlying shifts in content. *At play in the fields of consciousness: Essays in the honour of Jerome L. Singer* (pp. 29–50). Mahwah, NJ: Erlbaum.

Kotsiantis S., Kanellopoulos D., & Pintelas P. (2006) Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering*, 30(1), 25–36.

Kubat M. & Matwin S. (1997) Addressing the curse of imbalanced training sets: One-sided selection. *Proceedings of the 14th International Conference on Machine Learning*, 179–186.

Kyllonen, P.C. & Christal, R.E. (1990) Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14, 389–433.

Kyllonen, P.C. & Stephens, D.L. (1990) Cognitive abilities as determinants of success in acquiring logic skill. *Learning & Individual Differences*, 2, 129–160.

Lambert, D. (1992) Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1–14.

Liversedge, S.P., Paterson, K.B., & Pickering, M.J. (1998) Eye movements and measures of reading time. In Underwood, G. (Ed.), Eye guidance in reading and scene perception (pp. 55–75). Amsterdam: Elsevier.

Loboda, T.D. (2009) Pegasus [computer software]. Pittsburgh, PA, USA.

Maguire, E.A. (2001) Neuroimaging studies of autobiographical event memory. *Philosophical Transactions of the Royal Society of London. Series B: Biological sciences*. 356, 1441–1451.

Maki, R.H. & Berry, S. (1984) Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Menmory, & Cognition*, 10, 663–679.

Mason, M.F., Norton, M.I., Van Horn, J.D., Wegner, D.M., Grafton, S.T., & Macrae, C.N. (2007) Wandering minds: The default network and stimulus-independent thought. *Science*, 315, 393–397.

McConkie, G.W., 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.

McConkie, G.W. & Rayner, K. (1976) Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society*, 8, 365–368.

McCullagh, P. & Nelder, J.A. Generalized Linear Models, Second Edition, London: Chapman & Hall.

McVay, J.C. & Kane, M.J. (2007) Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 35(1), 196–204.

Miyake, A., Just, M.A., Carpenter, P.A. (1994) Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language, 33*, 175–202.

Morris, R.K., Rayner, K., & Pollatsek, A. (1990) Eye movement guidance in reading: The role of parafoveal letter and space information. *Journal of Experimental Psychology: Human Perception & Performance*, 16, 268–281.

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

Nelder, J.A. & Wedderburn, R.W.M. (1972) Generalized linear models. *Journal of the Royal Statistical Society*, Series A, General, 135, 370–384.

Nuthmann, A. (2005) The "where" and "when" of eye fixations in reading. PhD Dissertation.

Nuthmann, A., Engbert, R., & Kliegl, R. (2005) Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision Research*, 45(17), 2201–2217.

O'Regan, J.K. (1979) Eye guidance in reading: Evidence for the linguistic control hypothesis. *Perception & Psychophysics*, 25, 501–509.

O'Regan, J.K. (1980) The control of saccade size fixation duration in reading: The limits of linguistic control. *Perception & Psychophysics*, 28, 112–117.

O'Regan, J.K. (1990) Eye movements and reading. *Eye movements and their role in visual and cognitive processes* (pp. 395–453). Amsterdam: Elsevier.

O'Regan, J.K. (1992) Optimal viewing position in words and the strategy-tactics theory of eye movements in reading. *Eye movements and visual cognition: Scene perception and reading* (pp. 333–354). New York: Springer-Verlag.

Piñeiroa, G., Perelmanb, S., Guerschmanb, J.P., & Paruelo, J.M. (2008) How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecological Modelling*, 216, 316-322.

Plutowski, M., Sakata, S., & White, H. (1994) Cross-validation estimates IMSE. In J.D. Cowan, G. Tesauro, & J. Alspector (Eds.) *Advances in Neural Information Processing Systems 6* (pp. 391-398). San Mateo, CA: Morgan Kaufman.

Posner, M.I. Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3-25 (1980)

Pylkkanen, L. (2007) Lecture notes: "Neural Bases of Language". *Retrieved February 10, 2010 from http://www.psych.nyu.edu/pylkkanen/Neural_Bases/07_slides/11_LexAccess_Elect.pdf*.

R Core Team (2010) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria (http://www.R-project.org).

Radach, R., & Heller, D. (2000) Relations between spatial and temporal aspects of eye movement control. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 165–191). Oxford, England: Elsevier.

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

Rayner, K. (1998) Eye movement in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, (3), 372–422.

Rayner, K. (2009) The Thirty Fifth Sir Frederick Bartlett Lecture: Eye movements and attention during reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*, 1457–1506.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. (2004) The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception & Performance*, 30, 720–732.

Rayner, K. & Bertera, J.H. (1979) Reading without a fovea. *Sience*, 206, 468–469.

Rayner, K. & Duffy, S.A. (1986) Lexical ambiguity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14, 191–201.

Rayner, K. & Fisher, M.H. (1996) Mindless reading revisited: Eye movements during reading and scanning are different. *Perception & Psychophysics*, 58, 734–747.

Rayner, K. & Frazier, L. (1989) Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 15, 779–790.

Rayner, K. & McConkie, G.W. (1976) What guides a reader's eye movements. *Vision Research*, 16, 829–837.

Rayner, K. & Morris, R.K. (1992) Eye movement control in reading: Evidence against semantic preprocessing. *Journal of Experimental Psychology: Human Perception & Performance*, 18, 163–172.

Rayner, K. & Pollatsek, A. (1987) Eye movements in reading: A tutorial review. In Coltheart, M. (Ed.), *Attention & performance*, 12, 327–362.

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

Rayner, K., Pollatsek, A., & Binder, K. (1998) Phonological codes and eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 24, 476–497.

Rayner, K., Pollatsek, A., Drieghe, D., Slattery, T.J. & Reichle, E.D. (2007) Tracking the Mind During Reading Via Eye Movements: Comments on Kliegl, Nuthmann, and Engbert (2006) *Journal of Experimental Psychology: General*, 136, 520–529.

Rayner, K. & Raney, G.E. (1996) Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review*, 3, 238–244.

Rayner, K. & Well, A.D. (1996) Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3, 504–509.

Rayner, K., Well, A.D., & Pollatsek, A. (1980) Asymmetry of the effective visual field in reading. *Perception & Psychophysics*, 27, 537–544.

Rayner, K., Well, A.D., Pollatsek, A., & Bertera, J.H. (1982) The availability of useful information to the right of fixation in reading. *Perception & Psychophysics*, 31, 537–550.

Rayner, K., White, S.J., Kambe, G., Miller, B., & Liversedge, S.P. (2003) On the processing of meaning from parafoveal vision during eye fixations in reading. In J. Hyönä, R., Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 213–234). Oxford, England: Elsevier.

Reichle, E.D. (2006) Computational models of eye-movement control during reading: Theories of the "eye-mind" link. *Cognitive Systems Research*, 7, 2–3.

Reichle, E.D. (2010) *Computational models of reading*. Oxford University Press. Manuscript in preparation.

Reichle, E.D., Pollatsek, A., Fisher, D.L., & Rayner, K. (1998) Toward a model of eye movement control in reading. *Psychological Review*, 105, 125–57.

Reichle, E.D., Rayner, K., & Pollatsek, A. (1999) Eye movement control in reading: Accounting for initial fixation locations and refixations within the E-Z Reader model. *Vision Research*, 39, 4403–411.

Reichle, E.D., Rayner, K., & Pollatsek, A. (2003) The E-Z Reader model of eyemovement control in reading: Comparisons to other models. *Behavioral & Brain Sciences*, 26, 445-476.

Reichle, E.D., Liversedge, S.P., Pollatsek, A., & Rayner, K. (2009) Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences*, 13, 115–119.

Reichle, E.D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher-level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, 16, 1–21.

Reichle, E.D., Reineberg, A.E., & Schooler, J.W. (2010) Eye movements during mindless reading. *Psychological Science*, 20(10), 1–11.

Reilly, R. & O'Regan, J.K. (1998) Eye movement control in reading: A simulation of some word targeting strategies. *Vision Research*, 38, 303–317.

Robinson, A. (1995) The story of writing: Alphabets, hieroglyphs, and pictograms. New York: Thames & Hudson.

Robertson, I.H., Manly, T., Andrade, J., Baddeley, B.T., & Yiend, J. (1997) "Oops!" Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, 35, 747–758.

Rosenthal, R. (1998) Covert communication in classrooms, clinics, and courtrooms. *Eye on Psi Chi*, 3(1), pp. 18–22.

Rosenthal, R. (2002) Covert communication in classrooms, clinics, courtrooms, and cubicles. *American Psychologist*, 57(11), 839–849.

Rosenthal, R. (2003) Covert communication in laboratories, classrooms, and the truly real world. *Current Directions in Psychological Science*, 12(5), 151–154.

Salvucci, D.D. (2000) An interactive model-based environment for eye movement protocol visualization and analysis. In *Proceedings of the Eye Tracking Research & Applications Symposium*, 57–63. New York: ACM Press.

SAS Institute Inc. (2008) SAS/STAT 9.2 User's Guide. Cary, NC: SAS Institute Inc.

Sayette, M.A., Reichle, E.D., & Schooler, J.W. (2009) Lost in the sauce: The effects of alcohol on mind-wandering. *Psychological Science*, 20(6), 747–752.

Sayette, M.A., Schooler, J.W., & Reichle, E.D. (2009) Out for a smoke: The impact of cigarette craving on zoning-out during reading. *Psychological Science*, 21(1), 26-30.

Schacter, D.L. (2001) The seven sins of memory: How the mind forgets and remembers. Boston: Houghton Mifflin.

Schad, D., Nuthmann, A., & Engbert, R. (2012) Your mind wanders weakly, your mind wanders deeply – Objective measures reveal mindless reading at different levels. *Cognition*, 125, 179–194.

Schooler, J.W. (2001) Discovering memories in the light of meta-awareness. *Journal of Aggression, Maltreatment and Tmuma*, 4, 105–136.

Schooler, J.W. (2002) Representing consciousness: Dissociations between consciousness and metaconsciousness. *Trends in Cognitive Science*, 6, 339–344.

Schooler, J.W., Reichle, E.D., & Halpern, D.V. (2004) Zoning-out while reading: Evidence for dissociations between experience and metaconsciousness. *Thinking and seeing: Visual metacognition in adults and children* (pp. 203–226). Cambridge, MA: MIT Press.

Schooler, J.W., McSpadden, M., Reichle, E.D., & Smallwood, J. (2009) Unnoticed nonsense: Mind-wandering can prevent people from realizing that they are reading gibberish. Manuscript submitted for review.

Schustack, M.W., Ehrlich, S.F., & Rayner, K. (1987) The complexity of contextual facilitation in reading: Local and global influences. *Journal of Memory & Language*, 26, 322–340.

Seibert, P.S., & Ellis, H.C. (1991) Irrelevant thoughts, emotional mood states and cognitive performance. *Memory & Cognition*, 5, 507–513.

Shaw G.A. & Giambra, L.M. (1993) Task-unrelated-thoughts of college students diagnosed as hyperactive in childhood. *Deveopmental Neuropsychology*, 9, 17–30.

Singer, J.L. (1966) Daydreaming. New York: Random House.

Singer, J.L. (1988) Sampling ongoing consciousness and emotional experience: Implications for health. In Horowitz, M.J. (ed.), Psychodynamics and cognition, 297-346, Chicago: University of Chicago Press.

Smallwood, J. (2011) Mind-wandering while reading: Attentional decoupling, mindless reading and the cascade model of inattention. *Language and Linguistics Compass*, 5(2), 63–77.

Smallwood, J., Baracaia, S.F., Lowe, M., & Obonsawin, M.C. (2003) Task unrelated thought whilst encoding information. *Consciousness & Cognition*, 12, 452–484.

Smallwood, J., McSpadden, M., & Schooler, J.W. (2008) When attention matters: The curious incident of the wandering mind. *Memory & Cognition*, 36(6), 1144–50.

Smallwood, J., Obonsawin, M.C., & Heim, S.D. (2003) Task-unrelated thought: The role of distributed processing. *Consciousness & Cognition*, 12, 169–189.

Smallwood, J., Davies, J.B., Heim, D., Finnigan, F., Sudberry, M.V., O'Connor, R.C., & Obonsawin, M.C. (2004) Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness & Cognition*, 13, 657–690.

Smallwood, J., O'Connor, R.C., Sudberry, M.V., & Ballantyre, C. (2004) The consequences of encoding information on the maintenance of internally generated images and thoughts: The role of meaning complexes. *Consciousness & Cognition*, 4, 789–820.

Smallwood, J. & Schooler, J.W. (2006) The restless mind. *Psychological Bulletin*, 132, 946–958.

Smallwood, J., O'Connor, R.C., Sudberry, M.V., & Obonsawin, M.C. (2007) Mind wandering and dysphoria. *Cognition & Emotion*, 21, 816–842.

Smallwood, J., Nind, L., & O'Connor, R.C. (2009) When is your head at? An exploration of the factors associated with the temporal focus of the wandering mind. *Conscious Cognition*, 18(1), 118–125.

Smilek, D., Carriere, J.S.A., & Cheyne, A. (2010) Out of mind, out of sight: Eye blinking as indicator and embodiment of mind wandering. *Psychological Science*, 21(6), 786–789.

Starr, M.S. & Rayner, K. (2001) Eye movements during reading: Some current controversies. *Trends in Cognitive Science*, 5, 156–163.

Suppes, P. (1990) Eye-movement models for arithmetic and reading performance. In Kowler, E. (Ed.), *Eye movements and their role in visual and cognitive processes* (pp. 455–477). Amsterdam: Elsevier.

Suppes, P. (1994) Stochastic models of reading. In Ygge, J. & Lennerstrand, G. (Eds.), *Eye movements in reading* (pp. 349–364). Oxford: Pergamon Press.

Teasdale, J.D., Dritschell, B.H., Taylor, M.J., Proctor, L., Lloyd, C.A., & Nimmo-Smith, I. (1995) Stimulus-independent-thought depends upon central executive resources. *Memory & Cognition*, 28, 551–559.

Thibadeau, R., Just, M.A., & Carpenter, P.A. (1982) A model of the time course and content of reading. *Cognitive Science*, 6, 157–203.

Tulving, E. (1985) Memory and consciousness. *Canadian Psychologist*, 26, 1–12

Turner, M.L. & Engle, R.W. (1989) Is working memory capacity task dependent? *Journal of Memory & Language*, 28, 127–154.

Underwood, G. (2005) Cognitive processess in eye guidance. Oxford University Press.

Unsworth, N., Schrock, J.C., & Engle, R.W. (2004) Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 30, 1302–1321.

Unsworth, N., Redick, T.S., Heitz, R.P., Broadway, J.M., & Engle, R.W. (2009) Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, 17, 635–654.

Vincent, J.L., Snyder, A.Z., Fox, M.D., Shannon, B.J., Andrews, J.R., Raichle, M.E., & Buckner, R.L. (2006) Coherent spontaneous activity identifies a hippocampal-parietal memory network. *Journal of Neurophysiology*, 96(6), 3517–3531.

Vitu, F., McConkie, G.W., Kerr, P., & O'Regan, J.K. (2001) Fixation location effects on fixation durations during reading: An inverted optimal viewing position effect. *Vision Research*, 41(25-26), 3513–3533.

Vitu, E., O'Regan, J.K., Inhoff, A.W., & Topolski, R. (1995) Mindless reading: Eye movement characteristics are similar in scanning letter strings and reading text. *Perception & Psychophysics*, 57, 352–364.

Wegner, D.M. (1997) Why the mind wanders. In J.D. Cohen & J.W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 295–315). Mahwah, NJ: Erlbaum.

Weiss, S.M. & Kulikowski, C.A. (1991) Computer Systems That Learn. Morgan Kaufmann.

Williams, R.S. & Morris, R.K. (2004) Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, 16, 312–339.

Yang, S.N. & McConkie, G.W. (2001) Eye movements during reading: A theory of saccade initiation time. *Vision Research*, 41,3567–3585.

Zukerman, I. & Albrecht, D.W. (2001) Predictive statistical models for user modeling. *User Modeling & User-Adapted Interaction*, 11(1-2), 5–18.

Appendices

A. Questionnaire 1 (Pre-experiment)

1. If you have read any of the following books, how long ago was it (leave blank if you haven't):
   (a) Animal Farm: _____
   (b) Crime and Punishment: _____
   (c) Dr. Jekyll & Mr. Hyde: _____
   (d) Don Quixote: _____
   (e) Foundation: _____
   (f) Gone with the Wind: _____
   (g) Harry Potter: _____
   (h) Moby Dick: _____
   (i) Robinson Crusoe: _____
   (j) Sense and Sensibility: _____
   (k) The Old Man and the Sea: _____
   (l) The Lord of the Rings: _____
   (m) War and Peace: _____
2. List up to three of your favorite books genres:
   (a) _____
   (b) _____
   (c) _____

B. Questionnaire 2 (Post-experiment)

1. Sex: __ M __ F
2. Age: ____
3. Education: _____
4. Degree pursued: _____
5. Major/minor: _____
6. Occupation: __ Student __ Other: _____
7. Verbal SAT score: _____

In questions 8-14, please answer with a number 1 to 7, where 1 corresponds to "not at all" and 7 to "extremely."

8. How interesting was the novel (1-7)? ____
9. How focused on the text would you say you were (1-7)? ____
10. How stressed are you today (1-7): ____
11. How tired are you today (1-7): ____
12. How preoccupied have you been you with (1-7):
    (a) School: ____
    (b) Work: ____ *(put 0 if you don't work)*
    (c) Personal problems: ____
13. How likely would you be to participate in a similar experiment again (1-7): ____
14. If you were craving something during the experiment (e.g., food, drink, or cigarette) what was it and how much did you want it (1-7):
    (a) _____ ____ *(put 0 if you did not crave)*
    (b) _____ ____ *(put 0 if you did not crave)*
15. If you have seen the "Sense and Sensibility" movie, how long ago was it: ____
16. What were you thinking about while zoning out:
    _____
    _____

C. Reading Span Task Instruction

PAGE 1:
Welcome and thank you for your participation!

This part of the experiment consist of doing two tasks, A and B, at the same time. Good performance on both is important. Before the real test starts, you will have a chance to first practice both of the tasks separately, then both of them at the same time. On-screen instruction will tell you what to do.

All instructions will be presented in blue while stimuli will be presented in black.

Please read all instructions carefully and if at any point you have questions, please ask the experimenter.

PAGE 2:

Task A focuses on retaining a series of letters in memory for later recall. When the recall prompt is presented, choose letters in the order in which they were presented. If you forgot a particular letter, hit the blank box. The recall prompt will also allow you to reset your answer and start over. After you provide the answer you will be presented with the results (green = correct, red = incorrect).

PAGE 3:
Task B focuses on judging if English language sentences makes sense. You will be presented with one sentence at a time and be able to choose between "Correct" and "Incorrect." Below are examples of correct and incorrect sentences.

John was asked to have a seat on a chair. (correct)
John was asked to have a seat on a marmalade. (incorrect)

Your overall accuracy will be displayed in red in the top-right corner of the recall prompt screen. It is imperative that you answer as QUICKLY and ACCURATELY as possible. For our purposes we will not be able to use data with accuracy below 85%. Please try not to fall below that threshold.

D. Reading Task Instruction

PAGE 1:
Imagine the following scenario: You have to read a chapter of text for some class, so you sit down and start to read. At some point during reading, you realize that you have no idea what you just read. That is, you realize that, not only were you not really thinking about the text, you were thinking about something else altogether.

Chances are, the above scenario sounds familiar to you. At some point, we have all sat down to read something (such as a textbook or a novel), but while we were reading we "zoned out." That is, our eyes keep moving and it felt like we were reading, but when we caught ourselves "zoning out," we suddenly realized that we had not been understanding the text for some time.

In this experiment, you will read first several chapters of the novel "Sense and Sensibility" for about 1.5 hours. You will read 2-3 chapters at a time. Each set of chapters consists of 10-15 pages. Read the novel as you would normally read any other book.

To move ahead in the text, press the "right" key; similarly, to move backwards in the text, press the "left" key. We suggest you let your right palm rest on the cursor keys while reading for easy navigation.

If at any point in time you catch yourself "zoning out," press the "Z" key. Occasionally, the computer will prompt you as to whether you are currently "zoning out." To answer yes, press the "Y" key, to answer no, press the "N" key.

After you have finished reading, there will be a short test of how well you understood what you read.

Press "Enter" to continue

PAGE 2:
It is possible you will hear some noise coming from outside this room. Try not to get distracted. If you need to go to the restroom, please wait until you are done with the current set of chapters.

It is VERY IMPORTANT that you refrain from changing the position of your head while reading and keep it in the position in which you performed calibration. If you must move, please try to wait until you finish the current chapter.

Also, we know you cannot see the experimenter because of the vertical bars next to your eyes and you may feel uncomfortable about being watched. However, the experimenter will be watching their screen and attend to you only once in a while to make sure you are sitting properly.

On the next screen you will see nine black dots. You will see this screen at the beginning and at the end of every set of chapters you will read today. When that screen is presented please look at each dot for about 1 second. The order in which you attend to the dots does not matter. When you are done press "right" to continue.

If you have any questions, please ask the experimenter now. Otherwise, press any key to start the experiment.