Pace Yourself: Intraindividual Variability in Context Use Revealed by Self-paced Event-related Brain Potentials

Brennan R. Payne and Kara D. Federmeier

Abstract

■ Event-related brain potentials (ERPs) have revealed multiple mechanisms by which contextual constraints impact language processing. At the same time, little work has examined the trial-to-trial dynamics of context use in the brain. In the current study, we probed intraindividual variability in behavioral and neural indices of context processing during reading. In a concurrent self-paced reading and ERP paradigm, participants read sentences that were either strongly or weakly constraining completed with an expected or unexpected target word. Our findings revealed substantial within-subject variability in behavioral and neural responses to contextual constraints. First, context-based amplitude reductions of the N400, a component linked to semantic memory access, were largest among trials eliciting the slowest RTs. Second, the RT distribution of unexpected words in strongly constraining contexts was positively skewed, reflecting an increased proportion of very slow RTs to trials that violated semantic predictions. Among those prediction-violating trials eliciting faster RTs, a late sustained anterior positivity was observed. However, among trials producing the differentially slowed RTs to prediction violations, we observed a markedly earlier effect of constraint in the form of an anterior N2, a component linked to conflict resolution and the cognitive control of behavior. The current study provides the first neurophysiological evidence for the direct role of cognitive control functions in the volitional control of reading. Collectively, our findings suggest that context use varies substantially within individual participants and that coregistering behavioral and neural indices of online sentence processing offers a window into these single-item dynamics. ■

INTRODUCTION

Language comprehension is fast and dynamic. Within less than a second after perceiving a word or utterance, a highly distributed set of neural systems is engaged in the service of decoding that sensory input into meaning and integrating it into an incremental message-level semantic representation (Federmeier, Kutas, & Dickson, 2015; Kutas & Federmeier, 2000). One way that a listener or reader may optimize language understanding given these temporal constraints is by making use of accumulating context information. As reviewed below, a large, long-standing literature attests to the impact of context on essentially every aspect of language processing and, moreover, has shown that there are multiple mechanisms at work (Federmeier, 2007). For example, context cannot only ease the integration of new information into the current message-level representation, but also, at least in some circumstances, afford the preactivation of anticipated upcoming linguistic information (Kuperberg & Jaeger, 2016; DeLong, Troyer, & Kutas, 2014; Kutas & Federmeier, 2000). However, there still remains much to learn about the precise nature of these mechanisms and, importantly, about when each of them can be (or tends to be) used. There is a growing electrophysiological literature that has begun to examine how the use of different language comprehension mechanisms varies as a function of factors such as task demands and individual differences (e.g., Wlotko & Federmeier, 2015; Federmeier, Kutas, & Schul, 2010; Federmeier, 2007). Yet, processing mechanisms can also vary within individuals performing the same task (e.g., across the two cerebral hemispheres; Federmeier, 2007), perhaps even on a moment-by-moment basis.

Indeed, there has been a recently growing interest in psycholinguistics on intraindividual variability in behavioral performance during language processing (e.g., Payne & Stine-Morrow, 2014; Staub & Benatar, 2013). A number of studies have illustrated that manipulations of linguistic complexity (at multiple levels of representation) not only affect mean RTs but also change the underlying shape of the RT distribution, such that some experimental conditions impact overall trial-to-trial variability or increase the proportion of trials eliciting extremely slow RTs, skewing the distribution (reviewed by Reingold, Sheridan, & Reichle, 2015). These findings illustrate that the language processing system does not always respond to linguistic difficulty in a stationary manner (i.e., identically across all trials). Rather, processing waxes and wanes over the course of an experiment within individual participants. Trials eliciting extremely slow RTs have been argued to index qualitatively different underlying cognitive processes, for example, reflecting cases wherein normal processing had been halted or disrupted (e.g., Staub & Benatar, 2013) or reflecting lapses of

University of Illinois at Urbana-Champaign

sustained attentional control (Luke & Henderson, 2013; McVay & Kane, 2012). Although increases in RT variability have been proposed as a behavioral biomarker of compromised neurological functioning (MacDonald, Nyberg, & Bäckman, 2006; Hultsch, MacDonald, Hunter, Levy-Bencheton, & Strauss, 2000), there exists almost no research delineating the neural processes that underlie trials eliciting extreme RTs in attentionally demanding tasks such as language comprehension.

To address this gap, in the current study, we used a concurrent self-paced reading and event-related brain potential (ERP) paradigm to examine how contextual constraints impact RT variability during reading and whether electrophysiological responses to contextual constraints vary as a function of RT variability. Prior ERP work has shown that timing is an important factor in sentence comprehension. For instance, the degree to which context impacts word processing depends in part upon presentation rate. The N400, a component linked to meaning processing and initial access to semantic memory (Kutas & Federmeier, 2011), shows delays at fast presentation rates (Kutas, 1987) as well as reduced sensitivity to sentential factors (Camblin, Ledoux, Boudewyn, Gordon, & Swaab, 2007; Swaab, Camblin, & Gordon, 2004) and reduced prediction-based facilitation (Wlotko & Federmeier, 2015). However, because traditional ERP language experiments use RSVP methods with an experimentally fixed presentation rate, it remains to be determined how context use is affected by timing differences that are under volitional control. Here, therefore, we propose the idea that readers can dynamically modulate their use of context information on a trial-to-trial basis, resulting in correlated variability in their controlled rate of input and their neural responses. We begin by reviewing in more detail what is known both about context effects on word processing, especially as revealed through ERPs, and about intraindividual variability in language processing, as revealed in behavior.

Contextual Constraints on Word Processing

It is uncontroversial that supportive contexts can be used to facilitate word processing during language comprehension. For example, lexical decision RTs are facilitated when preceded by a meaningful semantic context (Schwanenflugel & LaCount, 1988; Stanovich, 1980), and both self-paced reading times (Smith & Levy, 2013) and eye fixation durations are faster to highly predictable words embedded in lexically constraining contexts (Rayner & Well, 1996; see Staub, 2015, for recent review). Since the 1980s, electrophysiological studies of language comprehension have shown that words embedded in congruent contexts elicit a reduced N400, reflecting eased semantic memory access (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980, for a review).

Although the facilitative effects of context on neural and behavioral indices of word processing are robust, de-

bate continues regarding the mechanisms by which contextual constraints are used in real time (Kuperberg & Jaeger, 2016; Staub, 2015; Federmeier, 2007). Facilitative effects of prior semantic context have been hypothesized to arise from multiple different sources, including passive associative priming (Traxler, Foss, Seely, Kaup, & Morris, 2000; but see Coulson, Federmeier, Van Petten, & Kutas, 2005), reduced difficulty with postlexical integration (Huettig & Mani, 2016; Hagoort, 2005), anticipatory preactivation of upcoming semantic features (Delong, Quante, & Kutas, 2014; Kutas & Federmeier, 2000; Federmeier & Kutas, 1999), and more direct or "strong" prediction of specific upcoming lexical items and their associated features (i.e., orthographical and phonological representations; Kuperberg & Jaeger, 2016; Lau, Holcomb, & Kuperberg, 2013; Luke & Christianson, 2012; Federmeier & Laszlo, 2009; DeLong, Urbach, & Kutas, 2005). An important difficulty in dissociating prediction and preactivation accounts of facilitation from more passive, integrative accounts arises because both theories predict facilitation of semantic processing in supportive contexts.

Thus, one way that researchers have explored the extent to which language comprehenders predict upcoming information is by examining whether there are processing consequences when readers encounter sensory information that is inconsistent with a prediction (Federmeier et al., 2007). Early behavioral work found little evidence for such prediction costs—increases in the processing of congruent information that is unexpected based on a constraining context—and these null results were taken as evidence that readers do not routinely predict upcoming lexical information (e.g., Frisson, Rayner, & Pickering, 2005; Schwanenflugel & LaCount, 1988; Forster, 1981). Although prediction costs have been inconsistently observed in the behavioral literature, more recent work using ERPs has revealed responses that are specifically elicited when comprehenders encounter congruent but unexpected words in strongly constraining contexts, in the form of a late, anteriorly distributed positivity following the N400 (Thornhill & Van Petten, 2012; Van Petten & Luka, 2012; Federmeier et al., 2007, 2010, for a review).

For example, Federmeier, Wlotko, Ocha-Dewald, and Kutas (2007) jointly manipulated lexical expectancy (low vs. high cloze probability) and sentential constraint (weak vs. strongly constraining sentence context), producing sentences such as (a)–(d):

(a) **Strongly Constraining, Expected:** Sam could not believe her story was *true*.
(b) **Strongly Constraining, Unexpected:** Sam could not believe her story was *published*.
(c) **Weakly Constraining, Expected:** I was impressed by how much she *knew*.
(d) **Weakly Constraining, Unexpected:** I was impressed by how much she *published*.

Critically, the lexically identical unexpected items (*b*) and (*d*) were matched for cloze probability (~0%) across the two levels of contextual constraint, so that any additional processing in (*b*) compared with (*d*) is likely driven by the cost of encountering an unlikely but plausible word in a context that is strongly predictive of a different word. Federmeier et al. (2007) found that N400 responses were graded in magnitude by each condition's average cloze probability ($a < c < b = d$). Importantly, conditions *b* and *d* did not differ in N400 activity. However, only strongly constraining but unexpected items (*b*) were shown to elicit a late anterior positivity following the N400. Federmeier and colleagues (2007) argued that this component likely reflects the increased resource demands needed to override or suppress the anticipated word and perhaps to revise the message-level representation following prediction violations. A number of other studies have reported a similar late anterior positivity (DeLong, Groppe, Urbach, & Kutas, 2012; Delong et al., 2014; Van Petten & Luka, 2012; Thornhill & Van Petten, 2012; Federmeier et al., 2010) and have shown that this anterior potential is dissociable from the centroposterior P600 response that is observed to words that are semantically/syntactically anomalous or violate specific event structures (DeLong et al., 2014; Van Petten & Luka, 2012; Kuperberg, 2007).

Although the current literature establishes the existence of additional neural activity engaged as a result of encountering prediction violations, the mechanisms underlying these indices of prediction error are not well understood. For instance, prediction-related effects are not always observed (see Van Petten & Luka, 2012, for a review), and the benefits and (prediction-related) costs related to context use appear to vary as a function of factors such as such as age (Wlotko, Federmeier, & Kutas, 2012), verbal fluency (Federmeier, et al., 2010; Federmeier, Melennan, Ochoa, & Kutas, 2002), and literacy skill (Huettig & Mishra, 2014). Over and above individual differences, predictive processing may be dependent upon task demands, such as presentation rate (Wlotko & Federmeier, 2015; Camblin et al., 2007). In addition to determining the factors that are necessary for eliciting prediction costs, it is critical to determine the functional role that such prediction costs play. For instance, it is unclear exactly what neurocognitive function or functions the anterior positivity indexes, with theories ranging from error-related learning or Bayesian updating (Kuperberg & Jaeger, 2016), to suppression of strong predictions (Federmeier et al., 2007), or postlexical repair and revision processes (DeLong et al., 2014). One way to begin to delineate the likely multiple mechanisms underlying these prediction costs is to examine the relationship between neural and behavioral indices of context use and how these vary within individual readers.

Almost no work has examined the trial-to-trial dynamics of context use in language processing. Although both facilitation and costs related to context processing may

exist on average, it is likely that these effects vary substantially within individuals. Kuperberg and Jaeger (2016) have recently argued that prediction in language comprehension is probabilistic in nature (rather than all or none), and others have demonstrated that prediction costs are graded in nature, varying in their strength with the degree of contextual congruity (DeLong et al., 2014). Moreover, it has been argued that revising semantic predictions may require the recruitment of attentional or executive resources (Payne, Stites, & Federmeier, 2016; DeLong et al., 2014).

Typical ERP studies employ an experimenter-controlled RSVP paradigm, in which the rate of input is held constant. However, a critical component of natural reading is that readers control the rate of input. Indeed reading times are highly variable from word to word, reflecting both attributes of the item and intrinsic sources of variability (cf. Payne & Stine-Morrow, 2014). To our knowledge, only one previous study has examined whether reliable ERP indices of language processing can be observed in RSVP paradigms wherein readers are allotted control over the rate of input (i.e., the self-paced reading paradigm). Ditman, Holcomb, and Kuperberg (2007) showed that the concurrent recording of ERPs during self-paced reading yielded reliable ERP effects of N400-eliciting pragmatic semantic violations and P600-eliciting morphosyntactic violations. Although this study provided an excellent proof of concept that ERPs can be reliably recorded while readers control their rate of input (see also Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011, for coregistration of fixation durations and ERPs during reading), the question of how ERP indices of language processing vary as a function of trial-to-trial behavioral fluctuations remains unanswered.

Intraindividual Variability in Language Comprehension

Implicit in focusing on mean changes in behavior or brain activity is the assumption that experimental manipulations that impact cognitive processing remain somewhat stable within an individual over a short period of time (typically at least the length of an experiment). Under this approach, variability across repeated measurements within an individual participant is treated as "noise" to be brought under statistical or experimental control. However, a growing body of research has demonstrated that intraindividual variability in behavior and brain functioning is substantial in magnitude (relative to between-person differences) across cognitive, sensory, and motor functioning and that such trial-to-trial variability can confer unique insight into cognitive processing over and above the mean (e.g., Dinstein, Heeger, & Behrmann, 2015; Grady & Garrett, 2014; Jackson, Balota, Duchek, & Head, 2012; McVay & Kane, 2012; Balota & Yap, 2011; Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007; Nesselroade & Salthouse, 2004; Hultsch et al., 2000).

In behavioral psycholinguistics, there is a small but growing literature beginning to examine intraindividual variability in online language processing, largely focusing on effects of experimental manipulations on underlying RT distributions (Payne & Stine-Morrow, 2014; Johnson, Staub, & Fleri, 2012; White & Staub, 2012; Staub, White, Drieghe, Hollway, & Rayner, 2010; Balota & Spieler, 1999). Importantly, RTs during sentence processing show a characteristic rightward skew, like most RT data. This "slow tail" is ignored in traditional analyses, despite the fact that different aspects of linguistic complexity may differentially affect not only the location of a distribution, but also its underlying shape. A number of researchers have advocated investigating RT variability by fitting parametric RT distributions to data (Balota & Yap, 2011; Balota, Yap, Cortese, & Watson, 2008). For example, RT distributions can be well captured by the ex-Gaussian distribution, a convolution of the Gaussian and exponential probability distributions. Two parameters are derived from the Gaussian component of the distribution: The central tendency (i.e., the location of the distribution) is reflected in μ, and variability in the modal portion of the distribution is captured by σ, the standard deviation. The rate parameter, τ, from the exponential component of the ex-Gaussian distribution reflects the degree of rightward slowing in the tail of the distribution. Importantly, conditions that impact μ result in a shift in the distribution, conditions that impact σ result in increased variability in the modal portion of the RT distribution, and conditions that impact τ result in a selective increase in very slow RTs. Increases in τ also increase the empirical variance of the RT distribution (as the empirical variance is constrained to equal $\sigma^2 + \tau^2$).

A number of studies have illustrated that experimental manipulations not only affect mean RTs but can also increase the variability and proportion of extreme RTs. For example, Staub et al. (2010) showed that word frequency influences both the location (μ) and the slow tail (τ) of the distribution of eye fixations in sentence reading (see Balota & Spieler, 1999, for similar findings in lexical decision tasks), signifying that frequency effects occurred on almost all trials, thus shifting the distribution, but that low-frequency words also resulted in a substantially increased proportion of extremely slow RTs, skewing the distribution. Moreover, a broader literature in attentional control has highlighted that task demands that impact τ reflect increases in attentional demand or disruptions of sustained attentional control (Jackson et al., 2012; McVay & Kane, 2012; Tse, Balota, Yap, Duchek, & McCabe, 2010; Schmiedek et al., 2007). Indeed, individual differences in τ are reliable across tasks and time and have been shown to share substantial variance with individual differences in factors such as age (Tse et al., 2010), working memory (Schmiedek et al., 2007), and risk for Alzheimer disease (Balota et al., 2010). Payne and Stine-Morrow (2014) recently showed that RT distribution parameters derived from sentence reading are strongly related to individual differences in visual sensory ability, verbal working memory capacity, and processing speed in younger and older adults. This raises the possibility that the mechanisms underlying increased behavioral variability or skewing in language processing may be similarly indicative of qualitatively different underlying processing mechanism (though not necessarily the same ones that have been identified in the sustained attention literature).

At the same time, criticism has been raised regarding the validity of assigning neurocognitive interpretations to RT distribution characteristics (Matzke & Wagenmakers, 2009; Balota et al., 2008). For example, Matzke and Wagenmakers (2009) showed that parametric decompositions of RT distributions do not yield parameters that directly map onto components of the diffusion-decision model, a successful cognitive model of two-alternative forced-choice RT (Ratcliff & McKoon, 2008). Indeed, because RT distribution fitting is necessarily descriptive (similar to mean RTs), additional converging evidence is necessary to reach substantive conclusions about the mechanisms underlying intraindividual behavioral variability (Balota et al., 2008). Despite existing theoretical arguments that qualitatively distinct cognitive processes are responsible for eliciting very slow RTs (in the tail of a RT distribution; Staub & Benatar, 2013; McVay & Kane, 2012; Schmiedek et al., 2007), there has been very little work examining the neural processes that underlie such trial-to-trial behavioral variability.

Although the study of intraindividual variability in language comprehension has focused on behavioral methods, there has been a growing interest outside of psycholinguistics in using neuroimaging methods to explore trial-to-trial neural dynamics underlying cognitive performance (e.g., DeLorme, Miyakoshi, Jung, & Makeig, 2015; Dinstein et al., 2015). The use of EEG/ERP methods is particularly well suited for examining brain signal variability, in part due to their excellent temporal resolution, which is necessary for the detection of reliable single-item fluctuations in activity (e.g., Payne, Lee, & Federmeier, 2015). In the current study, we explored how volitional control over the rate of input during reading modulates intraindividual variability in context use and prediction. To do this, we examined the within-subject trial-to-trial dynamics of context use in sentence comprehension in a concurrent self-paced reading and ERP paradigm (cf. Ditman et al., 2007).

METHODS

Participants

Twenty-eight adults (11 women, mean age = 20 years, range = 18–35) from the University of Illinois at Urbana-Champaign community participated in the experiment for course credit. All were right-handed native speakers of English as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971), reported near 20/20 corrected or uncorrected vision, and had no prior history of neurological or psychiatric issues.

Materials

Sentences were 282 sentence frames, taken from Federmeier and colleagues (2007), half of which were strongly constraining and half of which were weakly constraining, with a sentence-final word that was either expected or unexpected but plausible. The unexpected items were lexically identical across constraint on the single-trial level, so that differences between these conditions can only be attributed to the prior sentence context. More information on item norming is presented in Federmeier et al. (2007). A follow-up sentence appeared after each critical sentence-final word, so that end-of-sentence reading times were not confounded with the end of the trial. This follow-up sentence was held constant across conditions.

EEG Recording and Processing

EEG was recorded from 26 evenly spaced silver–silver chloride electrodes embedded in an EasyCap (Electro-Cap, Inc., Eaton OH; following the same montage as in Federmeier et al., 2007). Electrodes were referenced online to the left mastoid and re-referenced offline to the average of the right and left mastoids. In addition, one electrode was placed on the left infraorbital ridge to monitor for vertical eye movements and blinks, and another two electrodes were placed on the outer canthus of each eye to monitor for horizontal eye movements. Electrode impedances were kept below 5 kΩ. The continuous EEG was amplified with a BrainAmpDC (Brain-Vision, LLC, Morrisville, NC) amplifier (bandwith filtered: 0.02–250 Hz) and recorded to hard disk at a sampling rate of 1000 Hz. The continuous EEG was further high-pass filtered offline (30 Hz), and epochs of EEG data were taken from 100 msec before stimulus onset to 1500 msec post-stimulus onset. Epochs were examined and marked for artifacts (drift, muscle activity, eye blinks, and eye movements). On average, a total of 5% ($SD = 6$%; range across participants ≤ 1–20%) of critical trials were marked as artifacts and not included in subsequent analyses. There were no reliable differences in artifact rates across conditions. One participant was dropped following artifact detection for an excessive number of artifacts (43% of trials), leaving $n = 27$ participants in the final analysis. ERPs were computed using the ERPLAB toolbox (Lopez-Calderon & Luck, 2014) with the EEGLAB toolbox (Delorme & Makeig, 2004). Statistical analyses on behavioral and electrophysiological data were carried out using the R language for statistical computing (R Development Core Team, 2015).

Procedure

Participants were seated 100 cm in front of a 21-in. CRT computer monitor. Stimuli were presented using Presentation experimental control software (Neurobehavioral Systems, Albany, CA). Each trial began with a warning sign (several plus signs on the screen) presented for 500 msec; the blank screen between the warning sign and the first word of the sentence varied randomly from 500 to 1200 msec (to prevent the consistent buildup of anticipatory slow-wave activity). Sentences were then presented word by word in the center of the screen. Each word was presented for a minimum duration of 100 msec, with an ISI of 300 msec. Thus, the minimum possible SOA was 400 msec. The overall SOA was determined by participants pressing the button to advance each word. A 3-sec pause separated each sentence. Hand used to advance the sentence was counterbalanced across participants to reduce systematic lateralized motor potentials in the averaged ERPs (cf. Ditman et al., 2007).

Participants were asked to minimize blinks, eye movements, and muscle movements while reading. They were instructed to read for comprehension and told that they would be asked questions about what they had read at the conclusion of the recording session. The recording session began with a short set of practice sentences to acclimate the participants to the task situation. The main experimental session was divided into four blocks of sentences, with participants taking a short rest between each block; recording time was approximately 1 hr. Participants could also take brief breaks between each trial as needed. After the recording session ended, participants completed a recognition test. As in Federmeier et al. (2007), a list of 240 words was selected such that, for each participant, 80 of the words were never seen as sentence-final words during the experiment and, of the remaining 160 words, 40 sentence-final words came from each experimental condition. Participants were asked to circle all the words that they remembered seeing as a final word of one of the sentences in the experiment. Delayed recognition memory accuracy was quantified via signal detection sensitivity using the nonparametric A-index (see Zhang & Mueller, 2005). Mean A was 0.77 (95% CI [0.67, 0.84]), indicating that participants were successfully discriminating between old and new words. Thus, participants appeared to be attending to the experimental materials.

RESULTS

Analyses are presented as follows. First, we report the ERP and RT results independently, following standard conventions for co-registration studies in sentence processing (Dimigen et al., 2009; Ditman et al., 2007). Grand-average ERPs are reported across all electrodes (following a direct replication of Federmeier et al., 2007). Behavioral analyses are then presented examining the effects of semantic expectancy and constraint on the entire RT distribution (cf. Balota et al., 2008). Critically, we then examine the within-subject coupling between behavioral and ERP indices of context processing by examining the degree to which trial-to-trial intraindividual RT variability was associated with changes in brain activity as reflected in ERPs. We used two methods to examine

the relationship between intratrial RT variability and ERPs: (1) a vincentile-sorted ERP method, in which ERPs are calculated contingent upon intrasubject and intracondition item RT quartiles (see below) and (2) RT-sorted single-trial ERP-image plots (Delorme et al., 2015), a method to visualize RT-dependent variability in single-trial event-related EEG activity.

ERPs

Figure 1 plots the grand-average ERP waveforms for expected and unexpected sentence-final target words in strongly and weakly constraining contexts at all 26 electrode sites. The position of the sites on the diagram approximates their position on the head, with anterior at the top and posterior at the bottom. Negative is plotted up in this and all subsequent figures.

First, it is important to note that the baseline period was equivalent across all conditions and no artifacts (motor or otherwise) from the concurrent self-paced reading task appeared to impact the ERP waveforms (cf. Ditman et al., 2007). Moreover, ERPs from the self-paced reading paradigm evoked characteristic brain responses for visual word processing in the RSVP paradigm, with clear sensory potentials observed over occipital (P1, N1 P2) and anterior (N1, P2) channels. Sensory potentials were followed by a large centroparietal negativity, the N400, which was graded by cloze probability, such that the largest N400 was elicited by unexpected items, followed by weakly constrained

expected items, and then strongly constrained expected items showed the greatest N400 facilitation, replicating Federmeier et al. (2007).

N400

Mean N400 amplitudes were measured within an a priori time window of 300–500 msec from eight centroparietal electrodes where N400 effects are typically largest (cf. Federmeier et al., 2007). A linear mixed-effects model was fit to the data with Expectancy and Constraint as fixed factors and subject and electrode as random effects. The model is analogous to repeated-measures ANOVA, with an additional random effect for channel site (see Payne et al., 2015). A maximal random-effects structure was fit across subjects and electrodes, excluding the correlations between random intercepts and slopes (Barr, 2013; Barr et al., 2013). A reliable Constraint × Expectancy interaction was found ($b = 1.52$ μV, $SE = 0.46$; log-likelihood ratio test: $\chi^2(1) = 9.49, p < .001$). Expected endings elicited N400s of smaller amplitude than unexpected endings within both strongly ($M_{SCU} = -1.21$ μV, $M_{SCE} = 1.77$ μV; $b = -3.28$ μV, $SE = 0.53, t = -6.13$) and weakly ($M_{WCU} = -0.93$ μV, $M_{WCE} = 0.83$ μV; $b = -1.76$ μV, $SE = 0.29, t = -6.03$) constraining sentence frames, replicating Federmeier et al. (2007). A direct comparison of the (lexically matched) unexpected items in the strongly and weakly constraining sentences revealed no significant difference between unexpected items

Figure 1. Grand-average ERP waveforms for expected and unexpected sentence-final target words in strongly and weakly constraining contexts at all 26 electrode sites. Between 300 and 500 msec, a centroposterior negativity (N400) is graded by cloze probability, with the largest N400 seen to unexpected items, followed by weakly constraining expected items and smallest for expected items, replicating Federmeier et al. (2007). Over anterior channels, a sustained negativity to strongly constrained expected items is observed, but the previously observed frontal positivity to strongly constraining but unexpected items is small or largely absent.
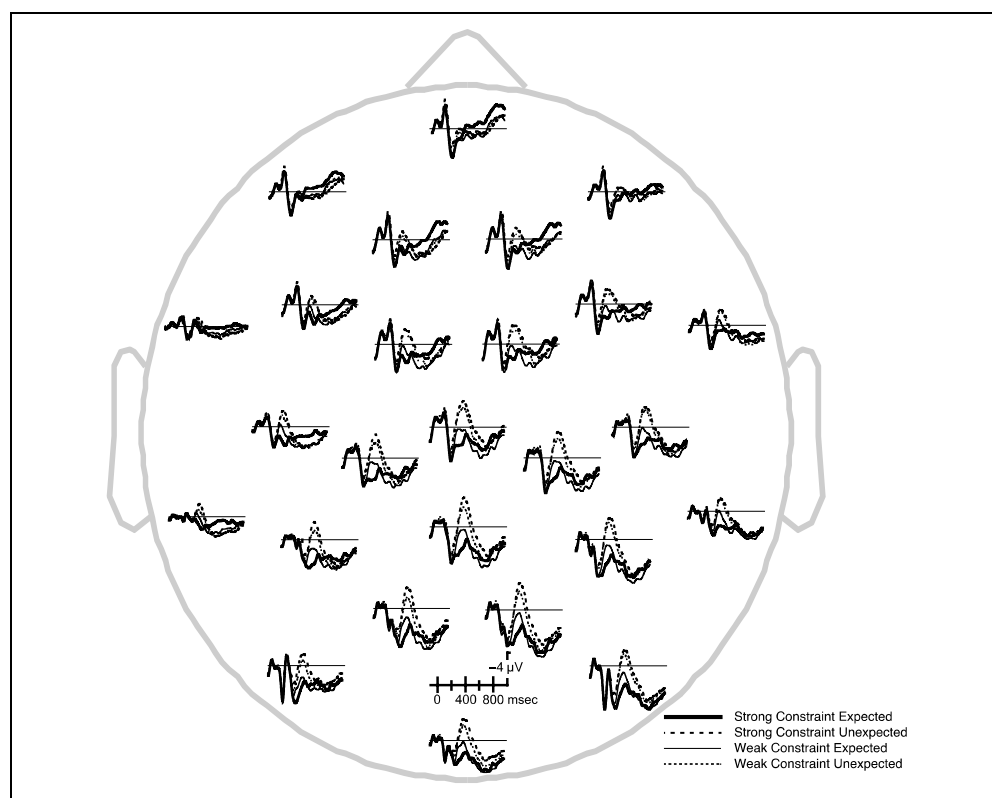
completing strongly and weakly constraining sentence frames ($b = 0.28$ μV, $SE = 0.24$; log likelihood ratio test: $\chi^2(1) = 1.34$, $p > .10$).

Late Anterior Potentials

Over anterior channels, a sustained negativity to strongly constrained but expected items was observed, as in Wlotko and Federmeier (2012), but notably the previously observed anterior positivity to strongly constraining but unexpected items (Wlotko & Federmeier, 2012; Federmeier et al., 2007) was either small or absent across anterior electrodes. To examine the effects of expectancy and constraint on the late anterior potentials, we conducted an analysis over the prefrontal and frontal electrode sites, within a time window of 500–900 msec, following Federmeier et al. (2007). Analyses on post-N400 activity over anterior channels revealed a reliable Constraint × Expectancy interaction ($b = 1.52$ μV, $SE = 0.46$; $\chi^2(1) = 6.39$, $p < .05$). Expected endings completing strongly constraining sentences elicited a more negative amplitude ($M_{SCE} = 0.001$ μV) relative to the weakly constraining but expected sentences ($M_{SCE} = 1.06$ μV; $b = -1.06$ μV, $SE = 0.43$, $t = -2.42$), consistent with prior work (e.g., Wlotko & Federmeier, 2012). A direct comparison of the unexpected items in the strongly and weakly constraining sentences revealed no reliable difference over anterior channels ($M_{SCU} = 0.84$ μV, $M_{WCU} = 0.95$ μV, $b = -0.11$, $SE = 0.29$, $\chi^2(1) = 0.16$, $p > .05$), consistent with the notable visual absence of the frontal positivity, which has been observed to unexpected items in strongly constraining contexts (Van Petten & Luka, 2012; Federmeier et al., 2007, 2010).

Reading Time Distributional Effects

Critical sentence-final reading times were trimmed within person and within condition for outliers using a conservative threshold at the 99.7th percentile. An ex-Gaussian distribution was then fit to the RT data separately for each participant and condition. Bootstrapped resampling ($n = 500$) was used to identify distribution parameters, which were estimated separately for each participant and condition via maximum likelihood estimation (see Van Zandt, 2000). The ex-Gaussian models converged normally for each participant.

Figure 2A presents the mean values for each of the ex-Gaussian parameters in each condition. For μ and σ, there were no reliable effects of target word expectancy or constraint (all ts < 1.50). However, as illustrated in Figure 2A, for strongly constraining but unexpected target words, the τ parameter was greater than all other conditions ($b_{SCE\ vs\ SCU} = 28$, $SE = 6.82$, $t = 4.15$; $b_{SCU\ vs\ WCE} = 22$, $SE = 6.82$, $t = 3.36$; $b_{SCU\ vs\ WCU} = 21$, $SE = 6.82$, $t = 3.10$), indicating that there was a larger proportion of extreme RTs when words were unexpected in contextually constraining sentence contexts. This can be readily observed in Figure 2B, which plots the probability density function of the estimated ex-Gaussian parameters as a function of constraint and expectancy. Density functions are estimated via Monte Carlo simulation (10,000 samples) from the model estimated ex-Gaussian parameters, where each distribution is generated by summing a sample from a normal distribution with mean μ and standard deviation σ and a sample from an exponential distribution with the rate parameter $1/\tau$. As can be seen, the underlying RT distributions are very similar across conditions. However, only unexpected words completing strongly constraining sentence contexts show a denser proportion of trials in the tail of the distribution, indicating a larger proportion of very slow RTs.

Finally, to assess distributional effects using a non-parametric method, vincentile plots, a descriptive method of examining RT distributions (Jiang, Rouder, & Speckman, 2004; Ratcliff, 1979), were constructed for each age group and each condition. The plots are constructed by rank-ordering RTs separately for each participant and condition and binning these RTs into quantiles (in this case, deciles) ranging from the fastest 10% of trials (vincentile 1) to the slowest 10% of trials (vincentile 10). Within-quantile averages across all participants are then plotted. This plot is presented in Figure 2C. Effects that are attributable only to an overall shift in the distribution (e.g., an effect on μ) would result in a condition difference that is equal across all vincentiles (i.e., parallel lines). In contrast, a shift in τ is observed when the conditions diverge at the largest vincentiles (i.e., the slowest trials). As can be seen, the plot confirms the findings from the parametric distributional fitting, with strongly constraining unexpected completions showing an increasing divergence from all other conditions at the slowest RTs.

To summarize the findings thus far, we observed clear prediction cost effects on RT behavior in the form of an increased proportion of slow RTs to unexpected words in strongly constraining contexts, thus skewing the RT distribution. However, in the grand-average ERPs, we did not observe such constraint effects on the anterior positivity, as is typically observed (Federmeier et al., 2007). The canonical approach in coregistration studies of sentence processing is to report and draw inferences about experimental influences on behavioral and neural data sources independently (e.g., Dimigen, Kliegel, & Sommer, 2012; Ditman et al., 2007). However, there are likely quantitative or qualitative differences in neural activity preceding trials eliciting comparably slow versus fast reading times, especially as a function of contextual constraint. That is, although we observed that prediction violations elicit an increased proportion of very slow reading times (see Figure 2), we have yet to observe whether the brain potentials corresponding to these trials vary, as typical ERPs are averaged across all trials. Therefore, as described in the following section, we examined the relationship between intraindividual variability in RT and their responses.
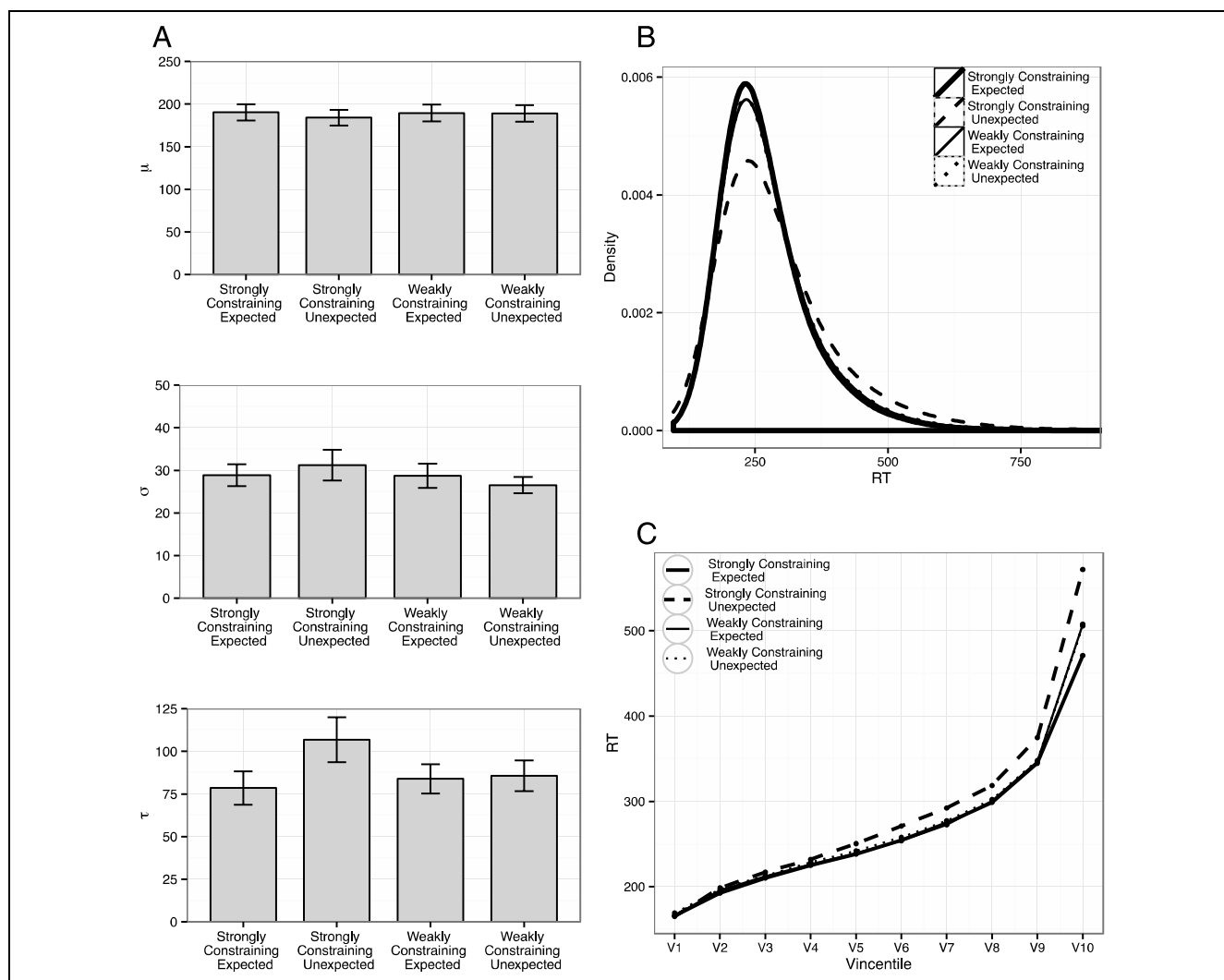
Figure 2. Behavioral results from analysis of RT distributions. A shows mean ex-Gaussian parameters (for expected and unexpected target words in strongly and weakly constraining contexts). B shows the corresponding simulated RT distributions, and C shows a vincentile plot of the empirical RTs (see text for details).

RT–ERP Coupling

To examine whether variability in context-related RT effects is coupled with qualitatively different electrophysiological responses, we examined the within-subject relationship between intraindividual variability in RTs and ERPs as a function of contextual constraint and expectancy. First, we conducted a modified RT-vincentile binning procedure (cf. Poli, Cinel, Citi, & Sepulveda, 2010), in which ERPs are back-sorted into separate bins based on RT quantiles. To strike a balance between the number of levels and the total number of trials per bin (for visualization of reliable ERP components), we grouped ERPs into lower, lower middle, upper middle, and upper quartiles based on single-word RT distributions. In contrast to prior reports using similar methods, whereby ERPs are binned by overall RT quantile (e.g., Jokeit & Makeig, 1994), we instead utilized a vincentile binning approach (see Reading Time Distributional Effects above) such that RTs were binned according to intrasubject and intra-

condition RT quantiles (i.e., quartile end-points were calculated separately for each participant and condition and used for calculating ERP averages). By binning within participant and condition RTs, this approach highlights ERP activity that is associated specifically with intrasubject and intracondition RT variability alone, rather than indexing overall participant response speed or overall condition influences. This approach also allows for a direct comparison with the behavioral analysis of the RT distribution. This approach results in an approximately balanced number of trials per bin (approximate only due to artifact rejection) and importantly provides a contribution of trials to bins that is balanced across participants and conditions.

N400

Figure 3A plots vincentile-contingent ERPs over a midline central site (corresponding to Cz in the 10–20 system), where N400 effects are typically largest. Visual inspection

of the ERPs revealed that there did not appear to be striking N400 differences in morphology or peak latency. All vincentile bins appeared to show some graded effect of target word expectancy on the N400, such that N400 amplitudes were greatest to completely unexpected completions, moderate to weakly constraining but expected completions, and smallest to strongly constraining but expected completions.

To formally test if the N400 effect latency varied across RT vincentiles, peak latency between 300 and 500 msec was measured on the unexpected conditions, which elicited the largest N400s. Because latency measures are nonlinear and exhibit a high degree of measurement error at the single-subject level, we adopted a jackknife approach to assessing peak latency (Kiesel, Miller, Jolicœur, & Brisson, 2008; Ulrich & Miller, 2001). The 50% fractional area latency (i.e., the time point before which 50% of the total area was observed) was measured separately for each jackknife sample from the Midline Central electrode (corresponding to Cz in the 10–20 system). There were no reliable differences in N400 latency across RT vincentiles (V1 = 399 msec, V2 = 403 msec,

V3 = 400 msec, V4 = 373 msec; $F_{corrected}$ = 0.37), indicating that the latency of the N400 response was largely invariant to intraindividual variability in reading times.

An omnibus test of the Constraint × Expectancy × Vincentile factor on mean amplitude was reliable ($\chi^2(3)$ = 13.40, $p < .001$). A graphical depiction of this interaction is plotted in Figure 3B. This interaction was largely driven by a decrease in N400 amplitude across RT vincentiles that was largest for expected completions in strongly constraining contexts, illustrating that there was greater contextual facilitation of the N400 among trials eliciting slower RTs. N400 amplitude to SCE completions were reliably larger for the fastest RT vincentile (V1) than all other vincentiles (b_{V1vsV2} = 1.60 μV, t = 6.95; b_{V1vsV3} = 1.93 μV, t = 8.40; b_{V1vsV4} = 1.55 μV, t = 6.74). Interestingly, although a similar effect was observed for the weakly constrained but expected words, this effect did not onset until trials in the third and fourth vincentiles (b_{V1vsV2} = −0.36 μV, t = −1.80; b_{V1vsV3} = 0.67 μV, t = 3.44; b_{V1vsV4} = 0.88 μV, t = 4.48). These findings suggest that trials eliciting the fastest RTs showed less contextual facilitation from the strongly constraining contexts than trials eliciting slower
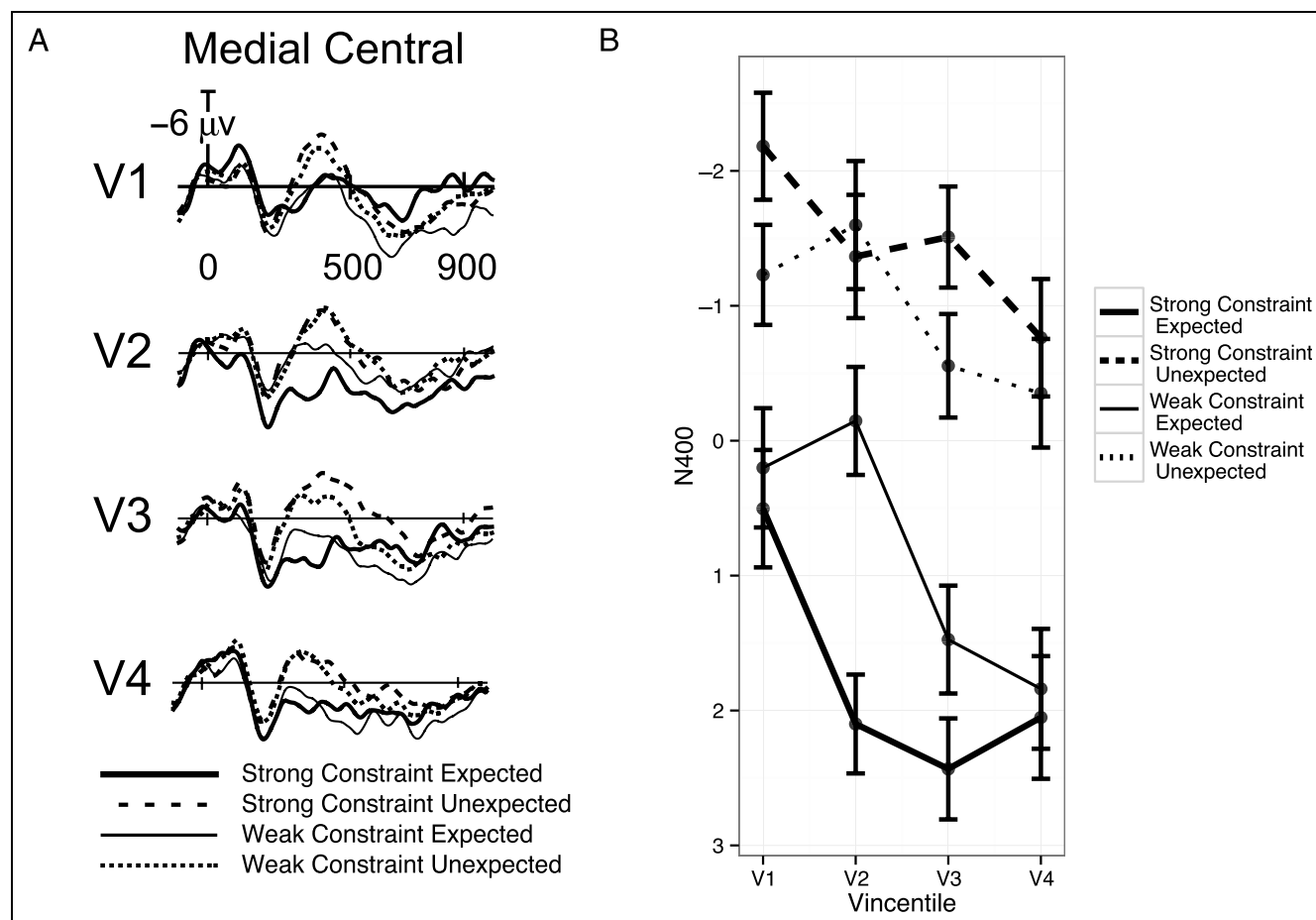
Figure 3. RT–ERP Coupling over centropartietal channels. A shows the ERP waveforms for the expected and unexpected final word in strongly and weakly constraining sentence contexts for each vincentile at a representative medial central electrode (corresponding to Cz in the 10–20 system). N400 amplitudes showed no clear latency modulation as a function of RT variability. However, trials eliciting the slowest RTs showed greater contextual facilitation, reflected in the amplitude reduction to the most constraining contexts. B shows mean N400 amplitudes (between 300 and 500 msec) averaged over the centroparietal electrodes used in the N400 statistical analysis.
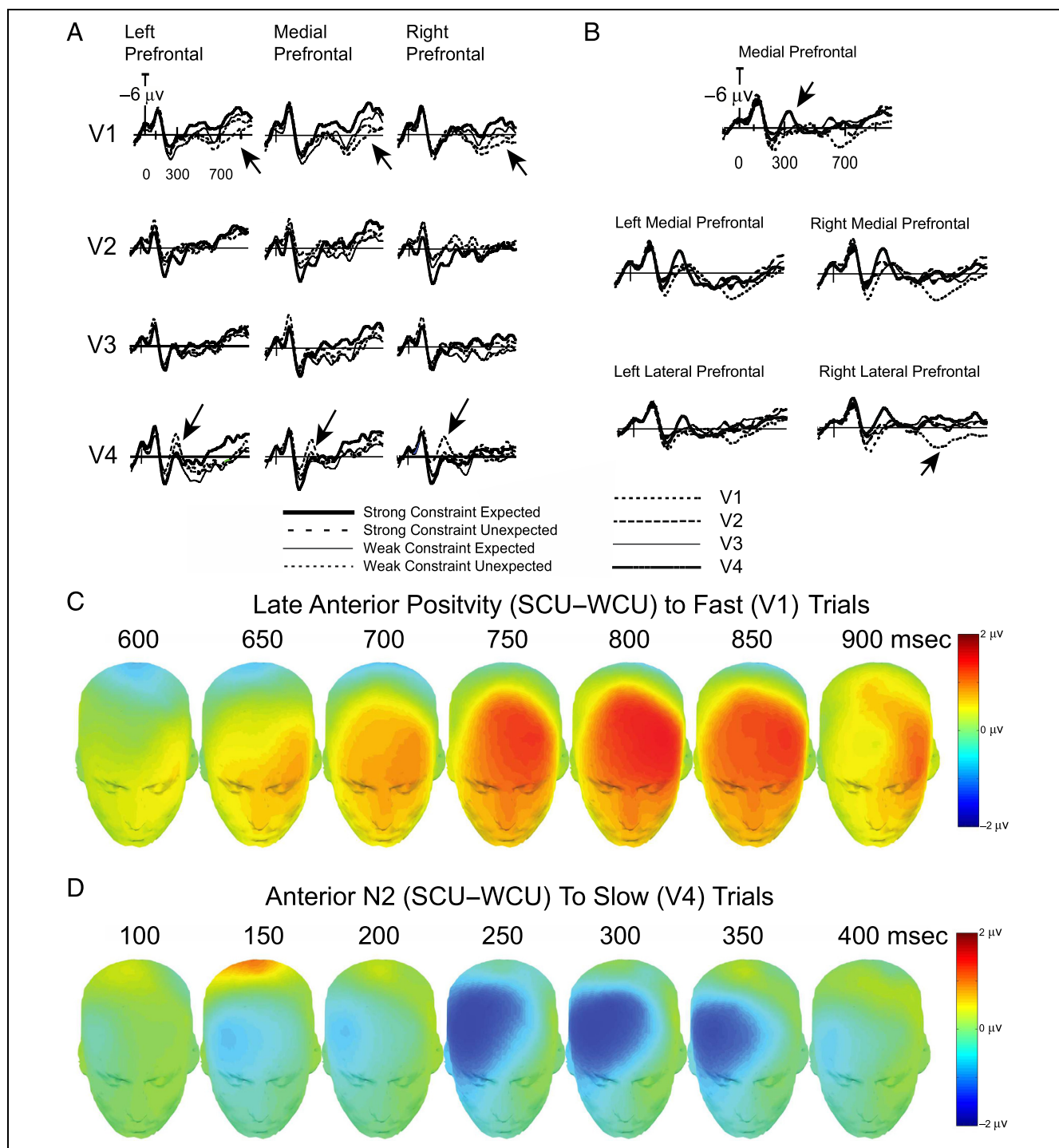
Figure 4. RT–ERP coupling over prefrontal channels. A shows ERPs to expected and unexpected sentence-final words in strongly and weakly constraining contexts over representative prefrontal channels, plotted separately for each RT vincentile. B plots vincentile-sorted ERPs to the strongly constraining but unexpected condition. Both plots illustrated that ERPs from trials eliciting fast RTs (V1) show the characteristic frontal positivity to strongly constraining but unexpected items. However, ERPs from trials eliciting slow RTs (V4) show an early anterior N2 to prediction violations. C plots the scalp topography of the late anterior positivity observed to prediction violations eliciting fast (V1) RTs, and D plots the scalp topography of the anterior N2 observed to prediction violations eliciting the slowest (V4) RTs.

RTs. Notably however, there were reliable expectancy effects across all RT bins, suggesting an intact N400 effect across all RTs (see also, Figure 5, bottom). Given the observation that constraint effects modulated the shape of the RT distribution such that prediction violations elic-

ited differentially skewed/slowed reading times, we then turned to focus on RT-dependent effects on the late anterior potentials, which are sensitive to semantic prediction violations (Van Petten & Luka, 2012; Federmeier et al., 2007).

Anterior Potentials

Figure 4A presents ERPs over representative medial and lateral anterior channels to unexpected and expected completions in strongly and weakly constraining sentences separately for each vincentile bin. Notably, the frontal positivity effect seen to strongly constraining but unexpected completions (Van Petten & Luka, 2012; Federmeier et al., 2007) is observed only for trials eliciting fast RTs (V1). Figure 4B provides a representation of the effect of RT on the magnitude of the anterior positivity by plotting the SCU completions across all vincentiles. This plot clearly illustrates the RT-dependent anterior positivity. Lastly, Figure 4C plots the topographical distribution of the anterior positivity effect (mean voltage of the SCU–WCU difference waveform) in consecutive 50 msec bins starting from the period following the N400. As can be seen, an anteriorly focused positivity to SCU items is observed between 700 and 900 msec, with a slightly left lateralization, consistent with prior investigations of the frontal positivity constraint effect (DeLong et al., 2014; Federmeier et al., 2007).

Although constraint effects were not observed on the late anterior positivity across slower vincentiles, in trials eliciting the slowest RTs (V4), a qualitatively different constraint effect emerged. Instead of eliciting a modulation of late (post-N400) anterior potentials, we observed a markedly early constraint effect in trials eliciting the slowest RTs, in the form of an anterior N2 potential to the strongly constraining but unexpected completions (see Figure 4A and C), peaking at approximately 300 msec. This anterior N2 constraint effect is particularly striking, as there was no clear evidence of an anterior N2 potential in the grand averages. Thus, by examining ERP activity contingent upon intraindividual RT variability, we discovered ERP dynamics that were otherwise obscured in the grand average. Figure 4D plots the scalp distribution of the anterior N2 constraint effect (SCU–WCU) across consecutive 50 msec bins from 100 to 400 msec. A more right lateralized anterior N2 was observed, with the strongest difference observed between 250 and 350 msec.

Frontal Positivity

An omnibus test of the Constraint $\times$ Expectancy $\times$ Vincentile factor was reliable ($\chi^2(3) = 40.12$, $p < .001$). To directly test whether constraint effects on the anterior positivity reliably vary as a function of intraindividual RT variability, a priori analyses focused on directly comparing the lexically matched unexpected items in strongly and weakly constraining sentence contexts across vincentiles. This analysis revealed a reliable Constraint $\times$ Vincentile interaction ($\chi^2(3) = 44.18$, $p < .001$) such that the magnitude of the frontal positivity effect was larger for V1 (1.12 µv; $t = 2.16$) than all other Vincentiles (*V2*: 0.28 µv; $t = 0.53$; *V3*: −0.70 µv; $t = 0.49$; *V4*: −0.27 µv; $t = 0.56$). Thus,

trials eliciting the fastest reading times showed the characteristic frontal positivity effect (Federmeier et al., 2007).

Anterior N2

Mean anterior N2 amplitudes were measured across the same 11 frontal and prefrontal channels as above across a typical anterior N2 latency band (200–400 msec). An omnibus test of the Constraint $\times$ Expectancy $\times$ Vincentile factor was reliable ($\chi^2(3) = 40.95$, $p < .001$). As above, analyses were focused on the contextual constraint effect by directly comparing the lexically matched unexpected items in the strongly and weakly constraining sentence contexts across vincentile bins. This analysis revealed a reliable Constraint $\times$ Vincentile interaction ($\chi^2(3) = 28.62$, $p < .001$) such that the magnitude of the constraint effect on the anterior N2 was largest for V4 (−1.19 µv; $t = −2.45$) and was absent in all other vincentiles (*V1*: −0.13 µv; $t = −0.27$; *V2*: −0.19 µv; $t = −0.33$; *V3*: −0.41 µv; $t = −1.12$).[1]

RT-sorted Single-trial ERP Images

We additionally adopted a single-trial approach to visualize the continuous relationship between individual RTs and ERPs. To visualize single-trial dynamics between RTs and ERPs, we created grand-average RT-sorted ERP images of the single-trial EEG activities (Delorme et al., 2015; Jung et al., 2001). Event-related single-trial EEG amplitudes are plotted over time, represented as color coded horizontal lines (as in spectral power in time–frequency representation). More negative amplitudes are plotted as colder colors and more positive amplitudes as warmer colors. Single trials are then sorted by RT (separately within condition and participant) and stacked vertically across items to form a 2-D colored image, where the *X*-axis represents time (as in ERPs), the *Y*-axis reflects trial, and the color-coding reflects signed amplitude. These images are computed separately for each participant and condition. To create "grand-average" ERP images for a condition, single-subject ERP images are then averaged. Because images are sorted by RT within participant and condition before averaging, the grand-average ERP image highlights event-related EEG activity as a function of intraindividual RT variability. That is, the ordering of trials and averaging of EEG activities are based on (participant and condition) relative RT, rather than absolute RT. Thus, visualizing the single-trial event-related EEG allows for an investigation into trial-level neural dynamics as they relate to RT variability.

Grand-average ERP images were computed following the methods described in Delorme et al. (2015). Figure 5 presents single-trial RT-sorted ERP image plots for the unexpected and expected targets in strongly constraining contexts over medial prefrontal (highlighting the anterior N2 and late positivity/negativity) and medial central
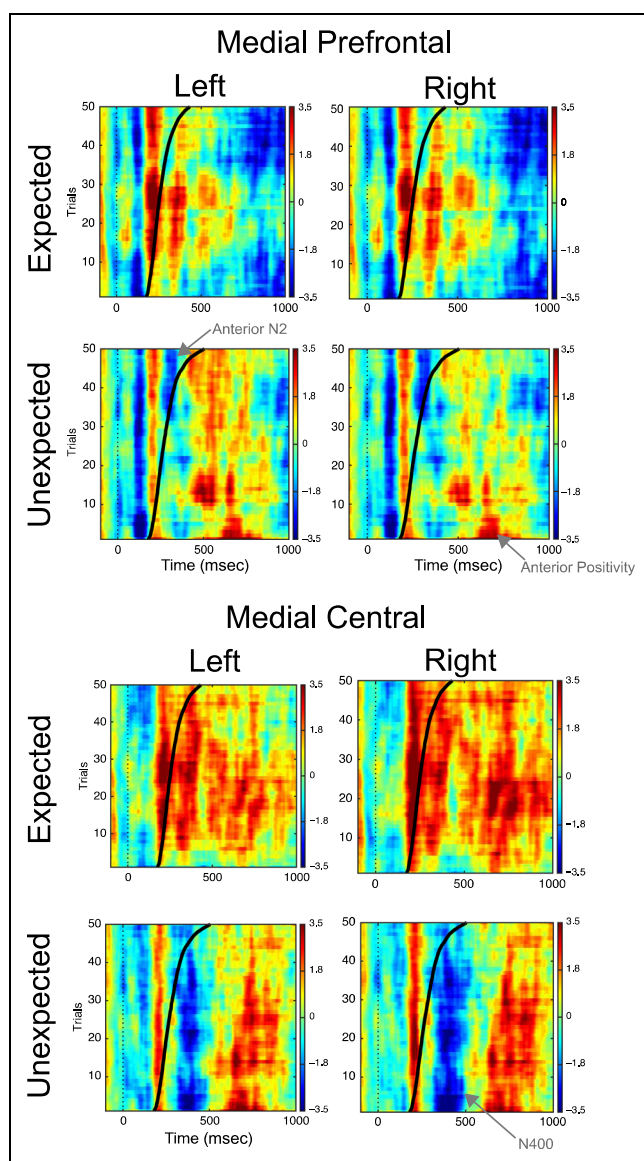
Figure 5. RT-sorted single-trial ERP images. The top panels show the grand-average ERP image for representative prefrontal sites for expected and unexpected target words completing strongly constraining contexts. The fastest trials elicit a late positivity to strongly constraining but unexpected items following the N400, but slow trials elicit an anterior N2. The bottom panel illustrates the same comparison over medial-central sites. See text for more details.

(highlighting N400 effects) channels. Single-trial RT values are superimposed on the ERP images as a solid black line, illustrating the sorting of trials by the distribution of RT values. Over the medial central electrode sites, a clear N400 is observed to strongly constraining unexpected completions, in the form of a relative negativity between 300 and 500 msec. Strikingly, although the amplitude of this potential varies somewhat from faster to slower RTs (with less activity at the slowest RTs), we see no evidence of the N400 showing RT alignment in latency. Thus, as clearly illustrated here, the N400 does not appear to vary in its temporal properties, even when

readers can control their own pacing. Additionally, comparing the activity in SCE, we can see that this same window shows a greater positivity (indicating N400 facilitation) as RTs increase.

Over the medial prefrontal channels, we observed the qualitatively different RT-dependent effects of constraint. For the fastest trials, a clear post-N400 positivity is observed approximately 500 msec poststimulus onset. This effect is reduced in amplitude for the intermediate and slowest RTs, however. The anterior N2 potential is clearly observed following the anterior N1/P2 complex. As can be seen clearly in this figure, whereas the N1/P2 complex appears across nearly every trial, the anterior N2 is observed only for those trials eliciting the slowest RTs. Indeed, this visualization appears to suggest a thresholding relationship between RTs and the anterior N2: Trials with RTs elicited before the peak of the anterior N2 show no clearly discernible anterior N2 activity, whereas trials with RTs elicited during or before the peak of the anterior N2 show strong anterior N2 activity, further signifying a tight coupling between the increased proportion of slowed RTs and the anterior N2 to contextual expectancy violations.

DISCUSSION

The current experiment was conducted to probe trial-to-trial variability in behavioral and neural indices of context use during reading. Our primary aim was to test the degree to which intraindividual variability in processing time, afforded by participants controlling the pacing of the input, was associated with variability in neural indices of context use. To examine this, we coregistered two of the most common online measures of sentence processing: self-paced reading time and ERPs. Toward this goal, we modified the traditional RSVP paradigm to allow readers to self-pace word by word (cf. Ditman et al., 2007) through sentences varying in semantic contextual constraint and the expectancy of the sentence-final target word. Our results clearly illustrated that contextual processing varied substantially from trial to trial in individual readers, as manifested jointly in both measures.

Notably, inspection of the reading time distributions revealed an increased skewing only among sentence-final words that violated strong semantic predictions. The finding that prediction violations only impacted the tail of the RT distribution (i.e., the exponential parameter τ) is noteworthy for a number of reasons. Researchers that have applied the ex-Gaussian distribution to eye fixation distributions during reading (Reingold et al., 2015; Staub & Benatar, 2013) have argued that factors that impact the tail of the RT distribution reflect an increase in the proportion of trials in which normal processing is disrupted. These are distinct from factors that uniformly shift an RT distribution (μ), which have been argued by some to index obligatory lexical processes (Reingold et al., 2015).

A small literature exists examining contextual influences on RT distributions during reading, largely focusing on the facilitative effects of contextual constraints (Staub & Benatar, 2013; Sheridan & Reingold, 2012; Staub, 2011). Staub had participants read the same set of target words twice, embedded in a highly constraining and a weakly constraining sentence context. They found that the distribution of fixation durations for words that are highly predictable was shifted to the left, impacting only the μ parameter, with no influences on τ. To our knowledge, our study is the first to examine the RT distributional effects associated with encountering an unexpected word in a strongly constraining context, and in contrast to prior contextual effects on RT distributions, we observed that semantic prediction violations did not elicit observable costs in behavior on all trials, but that for a subset of prediction violations reading was substantially slowed.

Effects on the τ parameter (and other exponential family parameters; e.g., Palmer, Horowitz, Torralba, & Wolfe, 2011) have been argued to index increases in response competition and other aspects of cognitively demanding post-perceptual/decision-related processing (McVay & Kane, 2012; Schmiedek et al., 2007; Balota & Spieler, 1999; but see Matzke & Wagenmakers, 2009), with increases in τ reliably occurring among conditions that are the most demanding. Moreover, individual differences in working memory capacity have been found to correlate quite strongly with τ in tasks that place high demands on attentional control (Tse et al., 2010; Schmiedek et al., 2007). The strong relationship between working memory capacity and τ was recently replicated in the sentence processing domain (Payne & Stine-Morrow, 2014), such that older adults and low WM span younger adults showed larger τ values (particularly at sentence boundaries) during self-paced reading. These findings suggest that trials in the tail of the RT distribution may index the subset of trials where attentional demands were greatest.

Both the "disrupted processing" and "increased attentional demand" accounts described above predict that trials in the tail of a RT distribution are functionally distinct from non-tail trials (in the Gaussian portion of the distribution). However, without additional converging evidence as to the nature of the processes that generate extremely slow RTs, a mechanistic account of the effects of prediction violations on trial-to-trial RT variability remains speculative. We therefore examined the ERPs underlying words that differed in their RT characteristics as a function of semantic expectancy and contextual constraints. Our RT–ERP coupling analyses revealed a number of important findings on the nature of intra-individual variability in context processing. We discuss these findings in two major parts. First, we discuss intra-individual variability in reading time and ERP indices of the benefits of context, primarily focusing on the N400 component. Second, we discuss how intraindividual variability in reading time is coupled with ERP indices of the

costs of context processing, focusing on semantic prediction violations.

First, we found that intertrial variability in reading time modulated the degree of contextual facilitation on the N400 component of the ERP. Although the N400 was largely graded with respect to semantic expectancy across all RT vincentiles, the degree of facilitation was dependent upon overall reading speed. For the most predictable words (SC-E), N400 amplitudes were reduced among trials that produced slower RTs (vincentiles 2–4), relative to trials that produced the fastest RTs (in vincentile 1). Thus, when readers allocated more time to use contextual constraints, this had a larger facilitative effect on semantic access for expected words. Interestingly, a similar pattern was observed for the moderately predictable endings (WC-E), although overall more time needed to be allocated before the N400 was reliably facilitated by these more weakly constraining sentences. Thus, despite the fact that the behavioral data did not show an overall effect of contextual facilitation, the coregistration results showed that RT did modulate the effects of semantic constraints on N400 activity. Indeed, these findings are consistent with prior work from the RSVP paradigm illustrating that (experimentally manipulated) longer presentation rates are associated with greater contextual facilitation on the N400 (Wlotko & Federmeier, 2015; Kutas, 1987). Notably, however, our findings suggest that, even when readers have control over their allocation of time, they do not always adjust their processing to maximize the use of contextual information (cf. Christianson, 2016; Ferreira & Patson, 2007).

Another striking finding from the current study was that the N400 showed remarkable temporal stability—latency was largely invariant of trial-to-trial RT variability. Indeed, latency invariance is a notable feature of the N400, with N400 latency only typically varying as a function of individual difference factors such as aging (Kutas & Iragui, 1998; but see Federmeier, Van Petten, Schwartz, & Kutas, 2003), dementia of the Alzheimer type (Olichney et al., 2002), and psychiatric disorders such as schizophrenia (Grillon, Ameli, & Glazer, 1991). Although stimulus factors such as familiarity, perceptual complexity, and frequency/probability impact the latency of some ERP components (e.g., the P3b; Luck & Hillyard, 1990), as well as RTs, N400 latency is rarely impacted by stimulus factors (although its amplitude and scalp distribution does vary reliably; see Kutas & Federmeier, 2011, for a review). Indeed, the temporal stability of the N400 has been theorized to serve a critical purpose in semantic memory access. Federmeier and Laszlo (2009) have argued that the temporal stability of the N400 affords the multimodal binding of stimulus features and knowledge representations to form the integrated representation of the meaning of an eliciting stimulus. In this sense, the timing of semantic access as indexed by the N400 is a critical feature of the dynamic construction of meaning representations during comprehension.

However, one factor that has been shown to modulate N400 latency is stimulus presentation rate. Kutas (1987) showed that, whereas slow to normal presentation rates (between 1–4 words per second) did not impact the latency of the N400, at very fast rates (~10 words per second), N400 amplitudes were disrupted, showing peak latency delays of over 75 msec. These findings suggest that very fast rates may disrupt the temporal synchronization that is critical to semantic access (Kutas & Federmeier, 2011). In the current study, although readers varied substantially in their self-paced reading times as a function of contextual constraints, N400 latency remained largely the same regardless of RT, likely because readers did not self-pace at a rate that was rapid enough to induce such N400 latency delays.

In contrast to the N400 findings, which largely reflected the benefits of context and appeared across nearly all trials (although modulated in strength as a function of RT variability), the costs associated with context processing—in the form of prediction violations—were highly variable from trial to trial in both behavior and ERPs. Prior ERP studies (DeLong et al., 2014; Federmeier et al., 2007, 2010) have shown that the brain often (but not always) elicits a positive potential following the N400 that is largest over prefrontal and frontal electrode sites when readers encounter an unexpected word in a strongly constraining sentence context (i.e., a context that allows for strong predictions). This activity appears to index the need to adjust processing when the incremental interpretation of text does not proceed normally because ones predictions' from the prior semantic context have been violated. Although the exact cognitive function(s) indexed by the frontal positivity are not yet fully understood, some have argued that these costs reflect the active inhibition of predicted-but-not-presented sentence completions and/or the revision or reinterpretation of a sentence context following prediction violations (see Van Petten & Luka, 2012, for a review), effects that have been argued by some to be attentionally demanding (Payne et al., 2016; DeLong et al., 2014).

Consistent with the claim that additional processes are brought online when unexpected words are encountered in the face of strong expectations for a different word, we found that RTs to these prediction violations were differentially skewed, suggesting an increased proportion of very slow RTs. However, the grand-average ERPs (see Figure 1) did not suggest evidence of a strong frontal positivity, as found in previous studies. Given this discrepancy between the electrophysiological and behavioral indices of prediction violations, we examined how ERPs varied as a function of intraindividual RT variability. In particular, we examined the brain activity underlying trials eliciting very fast reading times (i.e., which did not show strong prediction costs on behavior) versus those trials in the tail of the distribution (which did elicit behavioral effects in response to prediction violations). In only trials eliciting the fastest RTs (trials in the first vin-

centile), we observed that unexpected words completing a strongly constraining context indeed did elicit the typical late frontal positivity that has been observed previously in RSVP (Federmeier et al., 2007). It is interesting to note that average RTs in this quartile are roughly on par with the fixed presentation rate used in RSVP studies (~2 words per second). This suggests that when readers are pacing quickly through text, within a range similar to experimentally controlled RSVP studies, their brain responses to prediction violations mimic those that are found when they are not controlling the rate of input at all. We propose that this post-N400 frontal positivity may index a late revision process that is undertaken after semantic access is well underway or complete, when readers must revise the semantic representation in memory.

The theories reviewed above that attribute trials in the tail of an RT distribution (τ) to qualitatively and functionally distinct processes predict that such trials should elicit qualitatively distinct ERPs, representing an increase in resource demand or an electrophysiological index of disrupted processing. Indeed, in the slowest RTs, where prediction violation effects directly impacted reading behavior, the underlying brain activity showed a qualitatively distinct effect of constraint. Trials eliciting the slowest RTs (vincentile 4) to prediction violations showed a markedly earlier effect of contextual constraint on a component called the anterior N2, a negative going wave between 200 and 350 msec with a largely right lateralized prefrontal/frontal distribution.

The anterior N2 is a well-studied electrophysiological component that has been strongly linked to domain-general, immediate cognitive control of action (e.g., cancelling a prepared response, resolving conflicting response tendencies, Nieuwenhuis, Yeung, Van Den Wildenberg, & Ridderinkhof, 2003; Gehring & Willoughby, 2002). In their extensive review of the eliciting conditions of the anterior N2, Folstein and Van Petten (2008) highlighted that the amplitude of the anterior N2 has been found to vary as a function of conflict resolution and the need for cognitive control across a number of paradigms (the Stroop task, go/no-go, the Eriksen flanker task). Importantly, the anterior N2 is found most frequently in tasks that specifically place high demands on strategic cognitive control and conflict resolution in response selection (e.g., selection under high response competition demands or response inhibition; Bruin & Wijers, 2002; Gehring, Coles, Meyer, & Donchim, 1995). Moreover, studies attempting to localize the anterior N2 have found its sources to be consistent with generators in medial-frontal cortex, most reliably in the ACC, a cortical substrate that is often implicated in attentional control functions such as error monitoring and response inhibition (Bekker, Kenemans, & Verbaten, 2005; Nieuwenhuis et al., 2003).

Taken together, our findings suggest that the anterior N2 is triggered to a subset of trials that create prediction violations for individual readers. In these trials, the N2 acts as a motor inhibitory signal to immediately slow

reading time. Because of the early time course of this effect, we propose that the anterior N2 acts to inhibit the prepotent motor response to move forward in the subset of cases wherein readers make strong predictions about specific lexical candidates and therefore can very quickly inhibit the motor signal to continue reading when the early-available bottom–up information is inconsistent and causes conflict within the system. This brain activity reflected in the anterior N2 thus generates the increased proportion of very slow RTs in this condition. Although prior studies eliciting an anterior N2 typically used tasks that created conflict between two possible responses, our data show that such conflict-related brain activity can be observed in a natural task—reading for comprehension—that does not overtly require making a decision between externally induced conflicting responses. Instead, the activity underlying the anterior N2 in the current study is generated by internally generated conflict between a presented stimulus and a strongly held prediction that was actually never observed. Thus, there does appear to be a functional significance to trials eliciting extreme RTs when faced with increasing conflict, suggesting a direct role of cognitive control functions in reading. Indeed, this study, to our knowledge, provides the first evidence for the direct role of cognitive control functions, as indexed by the anterior N2, in the volitional control of reading, providing converging neurophysiological evidence that distinct neurocognitive processes are implicated in τ augmenting trials.

Collectively, the set of findings from the current experiment has important implications for understanding the dynamic range of strategies that may be continuously and probabilistically employed during sentence comprehension, particularly with regard to prediction and context use in reading. Moreover, this study highlights the utility of combining ERPs with trial-by-trial behavioral responses, providing a bridge between the growing behavioral literature on trial-by-trial variability in effects of linguistic constraints and the substantial electrophysiological literature on language processing. From the perspective of electrophysiological research, the current study demonstrates the importance of volitional control of input in modulating the neural response to contextual constraints as well as highlighting the substantial intraindividual variability in such responses—effects that were previously obscured in grand averages. Our RT–ERP contingent analysis revealed several ERP phenomena that were obscured in the grand average, notably the frontal positivity and anterior N2 effects to prediction violations. Thus, coregistration studies that typically only present the behavioral and neural results as independent signals (e.g., Dimigen et al., 2011; Ditman et al., 2007) should consider their conjoint relationship. Indeed, the results of the coregistration analyses call into question some fundamental assumptions regarding timing in the behavioral and eye-tracking literature on sentence processing. The current study showed that trials eliciting the slowest overall RTs (responses to prediction violations in the tail of the RT distribution) actually

showed the earliest effects of contextual constraints in the underlying brain activity as revealed by ERPs. To the extent that slowed behavioral responses as a function of conflict-resolution may derive from early domain-general cognitive control mechanisms, as indexed in the current study by the anterior N2, these findings have important implications for studies that derive the time-course of linguistic influences on word processing through behavioral responses alone (e.g., Rayner & Reingold, 2015). Lastly, from the behavioral perspective, the current study provides important inroads into the neurocognitive mechanisms that underlie trials eliciting different effects across the RT distribution, with our study showing the first evidence of increases in neural indices of cognitive control among trials eliciting extremely slow RTs to semantic conflict. Thus, the current study provides direct neurophysiological evidence that domain-general cognitive control functions are recruited in the volitional control of reading.

Acknowledgments

Note

1. To determine if individual differences in overall reading rate modulate the pattern of intraindividual variability in RT–ERP coupling, the vincentile models predicting mean N400, late anterior positivity, and anterior N2 amplitudes were re-fit with overall participant reading speed as a subject-level covariate. The inclusion of this covariate did not alter the pattern of results, suggesting that overall reading rate does not mediate the critical intraindividual variability findings.

REFERENCES

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128,* 32.

Balota, D. A., Tse, C. S., Hutchison, K. A., Spieler, D. H., Duchek, J. M., & Morris, J. C. (2010). Predicting conversion to dementia of the Alzheimer's type in a healthy control sample: The power of errors in Stroop color naming. *Psychology and Aging, 25,* 208.

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry the power of response time distributional analyses. *Current Directions in Psychological Science, 20,* 160–166.

Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language, 59,* 495–523.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328–329.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bekker, E. M., Kenemans, J. L., & Verbaten, M. N. (2005). Source analysis of the N2 in a cued go/nogo task. *Cognitive Brain Research, 22,* 221–231.

Bruin, K. J., & Wijers, A. A. (2002). Inhibition, response mode, and stimulus probability: A comparative event-related potential study. *Clinical Neurophysiology, 113,* 1172–1182.

Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C., & Swaab, T. Y. (2007). Processing new and repeated names: Effects of coreference on repetition priming with speech and fast RSVP. *Brain Research, 1146,* 172–184.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good enough, underspecified, or shallow language processing. *Quarterly Journal of Experimental Psychology, 69,* 1–29.

Coulson, S., Federmeier, K. D., Van Petten, C., & Kutas, M. (2005). Right hemisphere sensitivity to word-and sentence-level context: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 129.

DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, *121*, 226–239.

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*, 150–162.

DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass, 8,* 631–645.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word preactivation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8,* 1117–1121.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.

Delorme, A., Miyakoshi, M., Jung, T. P., & Makeig, S. (2015). Grand average ERP-image plotting and statistics: A method for comparing variability in event-related single-trial EEG activities across subjects and conditions. *Journal of Neuroscience Methods*, *250*, 3–6.

Dimigen, O., Kliegl, R., & Sommer, W. (2012). Trans-saccadic parafoveal preview benefits in fluent reading: A study with fixation-related brain potentials. *Neuroimage*, *62*, 381–393.

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

Dinstein, I., Heeger, D. J., & Behrmann, M. (2015). Neural variability: Friend or foe? *Trends in Cognitive Sciences, 19,* 322–328.

Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2007). An investigation of concurrent ERP and self-paced reading methodologies. *Psychophysiology, 44,* 927–935.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44,* 491–505.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41,* 469–495.

Federmeier, K. D., Kutas, M., & Dickson, D. S. (2015). A common neural progression to meaning in about a third of a second. In G. S. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 557–568). Holland: Elsevier.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language, 115,* 149–161.

Federmeier, K. D., & Laszlo, S. (2009). Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. *Psychology of Learning and Motivation, 51,* 1–44.

Federmeier, K. D., McLennan, D. B., Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*, 133–146.

Federmeier, K. D., Van Petten, C., Schwartz, T. J., & Kutas, M. (2003). Sounds, words, sentences: Age-related changes across levels of language processing. *Psychology and Aging, 18,* 858.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84.

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass, 1,* 71–83.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45,* 152–170.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology, 33,* 465–495.

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 862–877.

Gehring, W. J., Coles, M. G., Meyer, D. E., & Donchin, E. (1995). A brain potential manifestation of error-related processing. *Electroencephalography and Clinical Neurophysiology, 44,* 261–272.

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science, 295,* 2279–2282.

Grady, C. L., & Garrett, D. D. (2014). Understanding variability in the BOLD signal and why it matters for aging. *Brain Imaging and Behavior, 8,* 274–283.

Grillon, C., Ameli, R., & Glazer, W. M. (1991). N400 and semantic categorization in schizophrenia. *Biological Psychiatry, 29,* 467–480.

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences, 9,* 416–423.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience, 31,* 19–31.

Huettig, F., & Mishra, R. K. (2014). How literacy acquisition affects the illiterate mind—A critical examination of theories and evidence. *Language and Linguistics Compass, 8,* 401–427.

Hultsch, D. F., MacDonald, S. W., Hunter, M. A., Levy-Bencheton, J., & Strauss, E. (2000). Intraindividual variability in cognitive performance in older adults: Comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology, 14,* 588–598.

Jackson, J. D., Balota, D. A., Duchek, J. M., & Head, D. (2012). White matter integrity and reaction time intraindividual

variability in healthy aging and early-stage Alzheimer disease. *Neuropsychologia, 50,* 357–366.

Jiang, Y., Rouder, J. N., & Speckman, P. L. (2004). A note on the sampling properties of the Vincentizing (quantile averaging) procedure. *Journal of Mathematical Psychology, 48,* 186–195.

Johnson, R. L., Staub, A., & Fleri, A. M. (2012). Distributional analysis of the transposed-letter neighborhood effect on naming latency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1773–1779.

Jokeit, H., & Makeig, S. (1994). Different event-related patterns of gamma-band power in brain waves of fast- and slow-reacting subjects. *Proceedings of the National Academy of Sciences, U.S.A., 91,* 6339–6343.

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2001). Analysis and visualization of single-trial event-related potentials. *Human Brain Mapping, 14,* 166–185.

Kiesel, A., Miller, J., Jolicœur, P., & Brisson, B. (2008). Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology, 45,* 250–274.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146,* 23–49.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. *Language, Cognition and Neuroscience, 31,* 32–59.

Kutas, M. (1987). Event-related brain potentials (ERPs) elicited during rapid serial visual presentation of congruous and incongruous sentences. *Electroencephalography and Clinical Neurophysiology, 40,* 406–411.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4,* 463–470.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology, 62,* 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207,* 203–205.

Kutas, M., & Iragui, V. (1998). The N400 in a semantic categorization task across 6 decades. *Electroencephalography and Clinical Neurophysiology, 108,* 456–471.

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience, 25,* 484–502.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8,* 1–8.

Luck, S. J., & Hillyard, S. A. (1990). Electrophysiological evidence for parallel and serial processing during visual search. *Perception & Psychophysics, 48,* 603–617.

Luke, S. G., & Christianson, K. (2012). Semantic predictability eliminates the transposed-letter effect. *Memory & Cognition, 40,* 628–641.

Luke, S. G., & Henderson, J. M. (2013). Oculomotor and cognitive control of eye movements in reading: Evidence from mindless reading. *Attention, Perception, & Psychophysics, 75,* 1230–1242.

MacDonald, S. W., Nyberg, L., & Bäckman, L. (2006). Intra-individual variability in behavior: Links to brain structure, neurotransmission and neuronal activity. *Trends in Neurosciences, 29,* 474–480.

Matzke, D., & Wagenmakers, E. J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review, 16,* 798–817.

McVay, J. C., & Kane, M. J. (2012). Drifting from slow to "d'oh!": Working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 525–549.

Nesselroade, J. R., & Salthouse, T. A. (2004). Methodological and theoretical implications of intraindividual variability in perceptual-motor performance. *Journals of Gerontology, Series B, Psychological Sciences and Social Sciences, 59,* P49–P55.

Nieuwenhuis, S., Yeung, N., Van Den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive, Affective, & Behavioral Neuroscience, 3,* 17–26.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97–113.

Olichney, J. M., Morris, S. K., Ochoa, C., Salmon, D. P., Thal, L. J., Kutas, M., et al. (2002). Abnormal verbal event related potentials in mild cognitive impairment and incipient Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry, 73,* 377–384.

Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search?. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 58.

Payne, B. R., Lee, C. L., & Federmeier, K. D. (2015). Revisiting the incremental effects of context on word processing: Evidence from single-word event-related brain potentials. *Psychophysiology, 52,* 1456–1469.

Payne, B. R., & Stine-Morrow, E. A. (2014). Adult age differences in wrap-up during sentence comprehension: Evidence from ex-Gaussian distributional analyses of reading time. *Psychology and Aging, 29,* 213.

Payne, B. R., Stites, M. C., & Federmeier, K. D. (2016). Out of the corner of my eye: Foveal semantic load modulates parafoveal processing during reading. *Journal of Experimental Psychology: Human Perception and Performance, 42,* 1839–1857.

Poli, R., Cinel, C., Citi, L., & Sepulveda, F. (2010). Reaction-time binning: A simple method for increasing the resolving power of ERP averages. *Psychophysiology, 47,* 467–485.

R Development Core Team. (2015). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 446.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20,* 873–922.

Rayner, K., & Reingold, E. M. (2015). Evidence for direct cognitive control of fixation durations during reading. *Current Opinion in Behavioral Sciences, 1,* 107–112.

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

Reingold, E. M., Sheridan, H., & Reichle, E. D. (2015). Direct lexical and nonlexical control of fixation durations in reading. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 261–276).

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H. M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations

to working memory and intelligence. *Journal of Experimental Psychology: General, 136,* 414.

Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 344.

Sheridan, H., & Reingold, E. M. (2012). The time course of contextual influences during lexical ambiguity resolution: Evidence from distributional analyses of fixation durations. *Memory & Cognition, 40,* 1122–1131.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128,* 302–319.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16,* 32–71.

Staub, A. (2011). The effect of lexical predictability on distributions of eye fixation durations. *Psychonomic Bulletin & Review, 18,* 371–376.

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass, 9,* 311–327.

Staub, A., & Benatar, A. (2013). Individual differences in fixation duration distributions in reading. *Psychonomic Bulletin & Review, 20,* 1304–1311.

Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1280–1293.

Swaab, T. Y., Camblin, C. C., & Gordon, P. C. (2004). Electrophysiological evidence for reversed lexical repetition effects in language processing. *Journal of Cognitive Neuroscience, 16,* 715–726.

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology, 83,* 382–392.

Traxler, M. J., Foss, D. J., Seely, R. E., Kaup, B., & Morris, R. K. (2000). Priming in sentence processing: Intralexical spreading activation, schemas, and situation models. *Journal of Psycholinguistic Research, 29,* 581–595.

Tse, C. S., Balota, D. A., Yap, M. J., Duchek, J. M., & McCabe, D. P. (2010). Effects of healthy aging and early stage dementia of the Alzheimer's type on components of response time distributions in three attention tasks. *Neuropsychology, 24,* 300–315.

Ulrich, R., & Miller, J. (2001). Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology, 38,* 816–827.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83,* 176–190.

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review, 7,* 424–465.

White, S. J., & Staub, A. (2012). The distribution of fixation durations during reading: Effects of stimulus quality. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 603–617.

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *Neuroimage, 62,* 356–366.

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex, 68,* 20–32.

Wlotko, E. W., Federmeier, K. D., & Kutas, M. (2012). To predict or not to predict: Age-related differences in the use of sentential context. *Psychology and Aging, 27,* 975–988.

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika, 70,* 203–212.