Title
Half-listening or zoned out? It’s about the same: the impact of attentional state on word processing in context

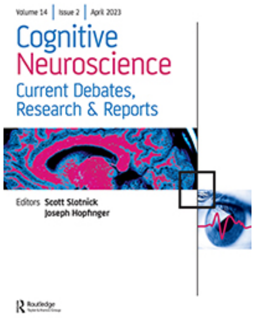
Journal
Cognitive Neuroscience, 14(3)

ISSN
1758-8928

Author
Boudewyn, Megan A

Publication Date
2023-07-03

DOI
10.1080/17588928.2023.2224959



Routledge
Taylor & Francis Group

Cognitive Neuroscience

Current Debates, Research & Reports

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pcns20

Half-listening or zoned out? It's about the same: the impact of attentional state on word processing in context

Megan A. Boudewyn

COGNITIVE NEUROSCIENCE
https://doi.org/10.1080/17588928.2023.2224959

Routledge
Taylor & Francis Group

Half-listening or zoned out? It's about the same: the impact of attentional state on word processing in context

Megan A. Boudewyn

Department of Psychology, University of California, Santa Cruz, CA, USA

ABSTRACT

Language comprehension must require some degree of attentional focus, but how do periods of inattention and/or split attention impact how language is processed? Here EEG was recorded while participants listened to full-length stories, and were periodically asked about whether they were fully attentive, were completely inattentive, or felt that they were in a split attention state. The ERP response to the words immediately preceding these attention questions was examined as a function of participant response, which allowed for the comparison of word processing in each of these attentional states. When participants were on-task, typical N400 effects of lexical frequency (smaller N400 for common compared to less common words), word position (smaller N400 for words appearing late in a sentence compared to words appearing with less preceding context), and surprisal (smaller N400 for relatively expected words compared to relatively unexpected words) were observed. When participants were in a fully inattentive state, the word-level effect of frequency was intact, but the context-dependent effects of word position and surprisal were significantly reduced. Interestingly, the pattern of results when participants were in a split attention state closely matched that of the fully inattentive state. Overall, the results demonstrate how attentional state influences sensitivity to language context during comprehension, and show that the consequences of inattention and split attention on word processing in context are quite similar, at least on the indices measured here.

ARTICLE HISTORY
Received 20 December 2022
Revised 2 May 2023
Published online 22 June 2023

KEYWORDS
ERPs; language; attention; comprehension

Introduction

Most of us engage in various forms of language comprehension all day, every day. We have conversations, read books and so on. However, our level of attention to what we are listening to or reading is not constant. Instead, our attention fluctuates. Sometimes our attention lapses completely, meaning that we temporarily fail to attend to the text or speech with which we were previously engaged. Other times we attempt to divide our attention, intentionally or unintentionally, between a language comprehension 'task' (like reading) and something else. Previous work shows that during language comprehension, our attention fluctuates relatively often: in language comprehension tasks conducted in a laboratory, readers and listeners self-report spending an average of 30–40% of time off-task (Boudewyn & Carter, 2018; Franklin et al., 2013).

Previous behavioral work has demonstrated that periods of inattention (but not split attention) during reading are predictive of poor performance on comprehension questions (Smallwood et al., 2008). Previous work has also used electrophysiology to investigate neural markers associated with attention lapses during language comprehension. In these studies, neural activity in the alpha frequency band was connected to periods of inattentiveness during language comprehension, as well as to poor comprehension performance for information presented during periods of particularly high levels of alpha activity (Boudewyn & Carter, 2018; Boudewyn et al., 2015). This data shed light on the neural markers associated with 'zone out states' as well as linked specific neural measures to comprehension outcomes.

However, an open question concerns how attentional state impacts lexical semantic processing of words in context. The identification and retrieval from memory of words as they are received during comprehension is influenced by a number of factors, including (but not limited to) how common or rare the word is (word frequency), whether the word occurs in isolation or after a substantial amount of context (word position), and how predictable or unpredictable the word is within that context (which can be quantified by a computationally-derived measure called surprisal). Previous work has also shown that an individual's ability

CONTACT Megan A. Boudewyn mboudewy@ucsc.edu Department of Psychology, University of California, Santa Cruz 1156 High Street, Santa Cruz, CA 95064, USA

to 'make use' of these various factors when processing incoming words varies as a function of individual differences in cognitive abilities such as working memory capacity, cognitive control and verbal skill (Boudewyn et al., 2012, 2013, 2015). *Within* an individual, however, does the influence of lexical characteristics and contextual information on word processing vary over time as a function of attention?

To address this question, the current study examined word processing in a story listening task during periods of inattentiveness and split attention. The primary focus of the analysis was on the N400 ERP component, a negative-going deflection elicited by meaningful semantic content (Kutas & Federmeier, 2011). The amplitude of the N400 is modulated by a variety of factors, including lexical characteristics (e.g., word frequency, such that more common words elicit a smaller N400 amplitude than less common words) and contextual information (e.g., predictability, such that more predictable words elicits a smaller N400 than less predictable words). Thus, the amplitude of the N400 is thought to reflect the relative ease of retrieval of a word's meaning from memory (Kutas & Federmeier, 2011).

Previous studies examining the influence of attention on the N400 have largely utilized word priming paradigms in which word pairs are presented at attended or unattended visual locations (Cristescu & Nobre, 2008; McCarthy & Nobre, 1993; Ruz & Nobre, 2008). For example, McCarthy and Nobre (1993) presented participants with word lists containing semantically related or repeated pairs of words, which were presented to either the left or right visual fields, and participants were instructed to attend left or attend right. Significant N400 effects of repetition and semantic priming were found for words in the attended visual field, but not in the unattended visual field (McCarthy & Nobre, 1993). However, in studies asking participants to direct visual attention to the location in which words are presented, preserved N400 effects of semantic relatedness have been found even without awareness, such as when words are presented so quickly after another visual stimulus that participants are not able to accurately report what they have read (Luck et al., 1996).

In the current study, participants listened to full-length stories while EEG was recorded, and were periodically asked about whether they were paying full attention to the task, were completely inattentive to the task, or felt that they were in a split attention state. ERPs to the content words in sentences immediately preceding the attention questions were examined as a function of participant response, allowing for the examination of word processing when participants were in each of these self-reported attentional states. It was hypothesized that when participants were fully attentive to the task, word processing would be significantly influenced by both lexical characteristics such as word frequency, and context-based factors such as word position and surprisal (a computationally-derived measure of a word's expectedness in context). In contrast, it was hypothesized that word processing would be less driven by such factors when listeners were in a split attention state, and least driven by such factors when listeners were in a fully inattentive state. In other words, a graded pattern of N400 effects as a function of attentional state was predicted, such that the largest effects would be observed when participants were on-task, and the smallest (or event absent) effects would be observed when participants were inattentive.

Methods

Participants

Sixty-six undergraduate students at the University of California, Davis received course credit for their participation in this experiment (44 female, 21 male, 1 decline-to-state; average age: 20.6, range 18–32 years). All were right-handed, native speakers of English, with no reported hearing loss or psychiatric/neurological disorders. Results from a separate analysis of oscillatory markers of attention lapsing in a 35-participant subset of this dataset have previously been published (Boudewyn & Carter, 2018). In addition, the data from 40 participants is available at https://osf.io/zft6e/as part of the supplemental materials for a recent paper in which a portion of the data was used to examine word representation vectors in natural language processing models (He et al., 2022).

Materials

Participants listened to two full-length short stories adapted from the Sherlock Holmes canon (*The Three Students* (Doyle, 1905) and *The Beryl Coronet* (Doyle, 1905)). Stories were adapted to conform to modern vocabulary and syntax norms for ease of listening (e.g., 'The Beryl Coronet' was modified to be titled 'The Emerald Crown'). Runtime for both stories combined was 72.8 minutes (*The Three Students*: 34.4 minutes; *The Emerald Crown*: 38.4 minutes). Stories were recorded with neutral intonation and speaking rate by a female speaker with a background in linguistics and previous experience recording language stimuli for similar experiments, and were digitally recorded using a Schoeps MK2 microphone (44,000 Hz, 16 bit).

At intervals throughout the story, attention probes appeared as text. A total of 54 probes over the course of both stories were presented (*The Three Students*: 25; *The Emerald Crown*: 29). The average time between probes 1.36 minutes for *The Three Students* (range: 0.78–2.13 minutes) and 1.39 minutes for *The Emerald Crown* (range: 0.47–2.21 minutes). A description with examples of each of the response categories to the attention probes ('on task,' 'off-task unaware' and 'off-task aware') was included in the task instructions. Probes read: 'Just prior to this question, was your attention on-task or off-task?' and given the following response options: 'on task,' 'Off task unaware (zoning out),' and 'Off task aware (tuning out).' 'on task' was defined to participants as an attentive state in which their attention was focused only on the task. 'Off task unaware (zoning out)' was defined to participants as an accidentally inattentive state, in which participants did not realize they were not attending to the task until the attention probe appeared. In contrast, 'Off task aware (tuning out)' was defined as a split attention state, in which participants may have felt they were partially attending to the task but were also knowingly thinking about something else. Participants were explicitly told that off-task thought is a typical occurrence during language comprehension, and were asked to respond to the attention probes truthfully based on their attentional state just before the probes interrupted the stories. The use of these attentional state categories and corresponding definitions was motivated by previous research on attention lapsing during language comprehension tasks in which significant differences have been observed between 'zoning out' and 'tuning out' states (Smallwood et al., 2008; Smallwood, McSpadden, et al., 2007). Thus, in the current study we operationalize an 'attentive' state using 'on task' responses, an 'inattentive' state using 'zoning out' responses, and a 'split attention' state using 'tuning out' responses.

Participants completed multiple choice questions about the stories after listening was complete. Please refer to Boudewyn and Carter (2018) for additional details on the task.

Content words in the two sentences immediately preceding each attention probe were marked as critical words for analysis, and their lexical frequency, word position, and surprisal were calculated.

Word Frequency was assessed using the 'SUBTL-WF' values for each word in the SUBTLEX-US word frequency database (Brysbaert & New, 2009). Low frequency critical words were defined as words with a frequency lower than 100 instances per million (average: 28.79; range: 0.02–98.92). High frequency critical words were defined as words with a frequency above than 100 instances per million (average: 355.84; range: 100.63–964.47).

Word Position was defined as the numbered position of a given critical word in a sentence. Early in Sentence words were defined as words in positions 2 through 6, and Late in Sentence words were defined as words in positions 7 through 12.

Surprisal was used to quantify the predictability of a given word in context. Surprisal is computed by taking the negative log transform of the probability of a given word in context. As such, high surprisal values for a word indicate that the word is relatively unpredictable in context, while low surprisal values indicate that a word is relatively predictable in context. There are many different ways to calculate the probability of a given word, including a variety of different language models, such as transformer language models and recurrent neural network models (Armeni et al., 2017; Frank & Willems, 2017; Frank et al., 2015; Kuperberg & Jaeger, 2016; Michaelov et al., 2021; Szewczyk & Federmeier, 2022). Here, we used the PsychFormers tool for calculating surprisal using transformer language models, which is available at https://github.com/jmichaelov/PsychFormers (Michaelov & Bergen, 2022). This tool was used to obtain surprisal values for all critical words using GPT−2, a transformer based language processing model (Radford et al., 2019). Previous work has found that predictability measures derived from GPT−2 covary with cloze probability (a measure of predictability derived from participant responses to a sentence completion task), and yield typical N400 effects of word predictability (Szewczyk & Federmeier, 2022) A median split of surprisal values was used to divide critical words into Relatively Predictable (average surprisal: 5.9; range: 0.03–9.99; average probability: 0.55; range: 0.5–0.87) and Relatively Unpredictable (average surprisal: 15.52; range: 11–41.56; average probability: 0.5; range: 0.5–0.5001) conditions.

It is worth noting that Word Position and Predictability were correlated, such that lower surprisal values (corresponding to more predictable words) were associated with higher word positions (corresponding to words later in the sentence) ($r = -0.35$). This is not surprising, as words might be expected to become more predictable in context as sentences unfold. However, these were separable variables, as only about 20% of critical words within the early Word Position condition overlapped with critical words in the relatively Unpredictable condition.

Procedure

After application of the EEG cap and facial electrodes, participants sat in a comfortable chair in an electrically-shielded, sound-attenuating booth. Participants were read task instructions by an experimenter, and then began the story listening task. During the task, participants listened to both stories (story order was counterbalanced across participants) and then completed a paper-and-pencil multiple choice comprehension test. During listening, a white fixation cross was presented in the center of a computer screen about 100 cm in front of the participants; when presented, attention probes were displayed as text on this screen.

EEG data acquisition and preprocessing

EEG was recorded using a 29-electrode custom Electro-Cap, with electrodes placed in standard positions according to the 10–20 system (ElectroCap International). To monitor for eye-movements and blinks, additional electrodes were placed on the outer canthi and below the left eye. The right mastoid was used as the recording reference, and the data was later re-referenced to the average of the left and right mastoids. EEG was recorded with a bandpass of 0.01 and 100 Hz, and digitized online at a sampling rate of 500 Hz. Impedances were kept below 5 kΩ during recording.

Preprocessing of EEG data was completed using the EEGLAB toolbox (Delorme & Makeig, 2004) with the ERPLAB plugin (Lopez-Calderon & Luck, 2014) for Matlab (https://www.mathworks.com/). ICA-based artifact correction was used to remove blink and other eye-movement artifacts in the data (Delorme & Makeig, 2004). After ICA was completed, the data was segmented into −200 to 1000 ms epochs time-locked to the onset of each critical word. Epochs were then screened for any remaining artifacts and baseline correction was performed using a −200 to 0 ms baseline window. A minimum of 20 artifact-free trials in all conditions was required for inclusion in the analysis; 13 participants were excluded based on this criterion. On average, 18% of trials were excluded following artifact rejection, and an average of 122 trials per condition were included in the final analysis. Artifact-free trials in each condition were averaged in order to create ERPs for each participant and condition. A low-pass filter with a 15 Hz half-amplitude cutoff was applied to all ERPs, and statistical analyses were conducted on the filtered data.

ERP analysis

Three comparisons of interest were tested as a function of attentional state: (1) Low Frequency vs. High Frequency words; (2) Early in Sentence vs. Late in Sentence words; and (3) Relatively Predictable vs. Relatively Unpredictable words. Analyses focused on 6 centro-parietal electrodes (CP1, CP2, P3, P4, Cz, and Pz) at which the N400 effect is typically maximal, and a single ERP was created for each participant and condition by averaging across these electrode sites. A repeated measures ANOVA was then conducted for each comparison of interest, using the mean amplitude in the 300–600 ms time window following the onset of a critical word at the dependent measure. Attentional State (3 levels: Attentive, Inattentive, Split Attention) and Trial Type (2 levels) were included as within-subjects factors. Significant interactions between Attentional State and Trial Type were followed-up with simple effects comparisons testing whether the effect of Trial Type was significant in each Attentional State (i.e., to determine whether the effect was present in each Attentional State). Where appropriate, an additional test was conducted to determine whether the effects of Trial Type were significantly different across Attentional States (i.e., to determine whether any significant effects of Trial Type were significantly larger for one Attentional State vs. another). This was done by dropping the factor of Trial Type from the rANOVA, and using the difference between Trial Type Level 1 and Trial Type Level 2 as the dependent measure.

Results

Word frequency effects

The omnibus rANOVA showed significant main effects of Word Frequency ($F(1,52) = 6.76$; $p = 0.012$) and Attention Status ($F(2,104) = 12.79$; $p = 0.001$), such that Low Frequency words elicited a larger N400 amplitude than High Frequency words, and larger N400 amplitudes were most negative when participants were Attentive. The interaction of Word Frequency and Attention Status was not significant. The effects of Word Frequency by Attention Status are plotted in Figure 1.

Word position effects

Significant main effects of Word Position ($F(1,52) = 44.47$; $p < 0.0001$) and Attention Status ($F(2,104) = 17.79$; $p < 0.0001$) were characterized by a significant interaction ($F(2,104) = 14.08$; $p < 0.0001$). Simple effects comparisons showed that the effect of Word Position was significant when participants were Attentive ($F(1,52) = 27.38$; $p < 0.0001$), Inattentive ($F(1,52) = 16.62$; $p = 0.00016$) and in a Split Attention

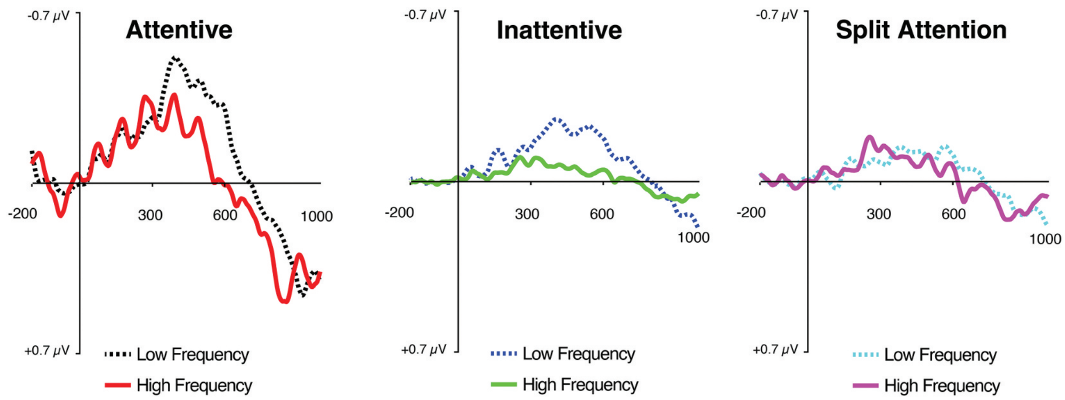


Figure 1. ERP effects of word frequency by attentional state. Plotted is the composite waveform at the 6 centro-parietal electrode sites used for all analysis (CP1, CP2, P3, P4, Cz, and Pz). Negative is plotted up.

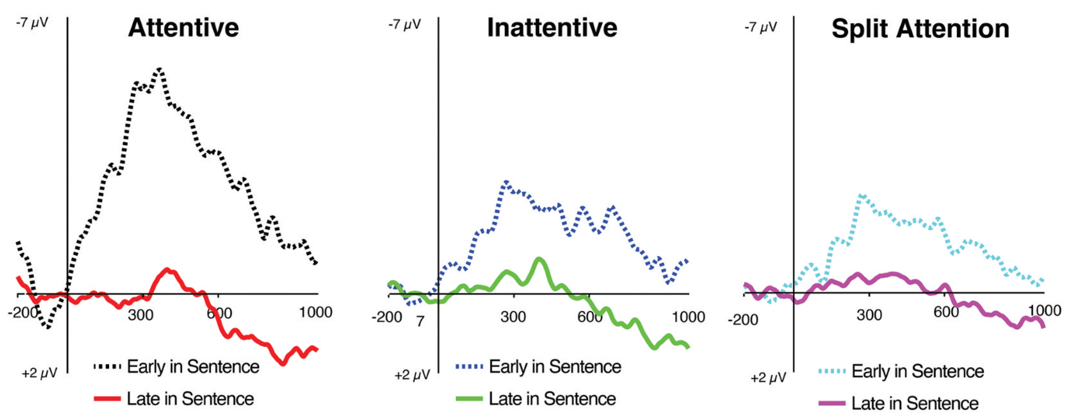


Figure 2. ERP effects of word position by attentional state. Plotted is the composite waveform at the 6 centro-parietal electrode sites used for all analysis (CP1, CP2, P3, P4, Cz, and Pz). Negative is plotted up.

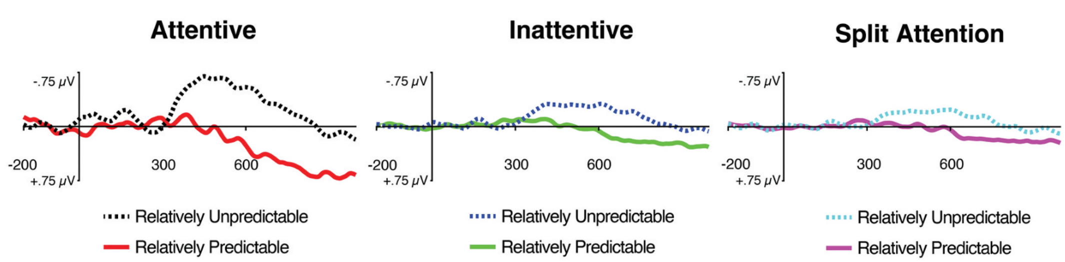


Figure 3. ERP effects of predictability by attentional state. Plotted is the composite waveform at the 6 centro-parietal electrode sites used for all analysis (CP1, CP2, P3, P4, Cz, and Pz). Negative is plotted up.

state ($F(1,52) = 18.13$; $p < 0.0001$). The effect of Word Position was significantly larger when participants were Attentive than when participants were in either an Inattentive state ($p = 0.0015$) or a Split Attention State ($p = 0.0011$). The Word Position effects for Inattentive and Split Attention States did not significantly differ from one another. To sum up: significant effects of Word Position were found in all attention conditions, but were significantly larger when participants were Attentive compared to off task. The effects of Word Position by Attention Status are plotted in Figure 2.

Predictability effects

In addition to significant main effects of Predictability ($F(1,52) = 9.67$; $p = 0.003$) and Attention Status ($F(2,104) = 9.68$; $p < 0.0001$), there was a significant interaction between the two ($F(2,104) = 3.68$; $p = 0.045$). Simple effects comparisons showed that the effect of

Predictability was significant when participants were Attentive ($F(1,52) = 7.56$; $p = 0.008$), Inattentive ($F(1,52) = 7.53$; $p = 0.008$) and in a Split Attention state ($F(1,52) = 6.52$; $p = 0.014$). The effect of Predictability was significantly larger when participants were Attentive than when participants were in a Split Attention state ($p = 0.039$); the effect when participants were in a Inattentive state did not significantly differ from either of the other two conditions. To sum up: significant effects of Predictability in the N400 window were found in all attentional states, and the effect was significantly larger when participants were Attentive than when they were in a state of Split Attention. The effects of predictability by Attention Status are plotted in Figure 3.

General discussion

This study examined word processing during story listening as a function of attentional state. The results showed that when participants confirmed that they were in an on-task, attentive state during listening, word processing was facilitated by both local word-level characteristics like lexical frequency and context-based factors like predictability (see Figure 4 for a summary of all attention effects). In other words, when they were paying attention, participants were able to quickly make use of a variety of contextual cues in order to facilitate the processing of incoming words. However, as might be expected, when participants were in an inattentive, 'zoned out' state, they were less sensitive to these contextual cues, showing reduced effects of context (although intact word-level effects). Interestingly, sensitivity to context was just as reduced (if not more so) when participants reported being in a state of split attention compared to fully inattentive.

This pattern of results provides insight into how language is processed (or partially processed) during a so-called 'zoned out' state, when participants reported being totally inattentive to the story. Some accounts of attention lapsing suggest that there is a 'perceptual decoupling' that occurs during an attention lapse, or period of mind-wandering, such that the locus of attention is shifted internally at the expense of attention to incoming external/perceptual input (Schooler et al., 2011; Smallwood, 2011; Smallwood, Fishman, et al., 2007). According to the cascade model of inattention, a shift of attention away from task-relevant stimuli will result in less detailed perceptual representations of those stimuli, which ultimately leads to more superficial processing than would occur when attention is engaged with the task (Smallwood, 2011). In line with this model, periods of inattention during language processing have been previously connected to reduced comprehension, which has been interpreted as indicative of more shallow processing (Smallwood et al., 2008). In the current study, evidence was found for relatively intact 'low level' (word level) processing during periods of inattention, whereas 'high level' processing (context-dependent effects) was significantly reduced. These results are consistent with the cascade model of inattention.

The pattern of results observed when participants were in a 'half-listening' or split-attention state was particularly intriguing. In contrast to the initial prediction that the results in this condition would be intermediate between the attentive and inattentive conditions, the results instead suggested that split attention is 'just as bad' as inattention, at least in terms of its effect on the language processing indices measured here. Anything other than a fully attentive state appeared to lead to relatively locally-driven processing, in that reduced effects of global context on word processing were found when either fully inattentive or dividing attention (compared to when fully on-task). This result is consistent with previous work that demonstrated attenuated N400 effects of semantic fit under dual task conditions in which attention was split between sentence reading and a sustained visual attention task (Hubbard & Federmeier, 2021). In addition, previously published analyses of a subset of the current data found that comprehension accuracy on a multiple-choice test completed after story listening was reduced to the same extent when participants were inattentive as when they reported being in

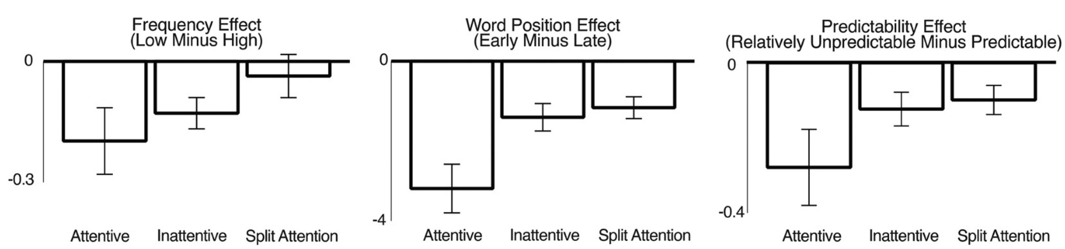


Figure 4. Summary of all N400 effects of attention. Values reflect the difference between conditions in mean amplitude between 300–600 ms for the composite waveform at the 6 centro-parietal electrode sites used for all analysis (CP1, CP2, P3, P4, Cz, and Pz). Note: as the N400 is a negative-going ERP, larger effects are indicated by more negative values.

a split attention state (Boudewyn & Carter, 2018). These results suggest that split attention is no better than experiencing a complete lapse of attention. However, some previous behavioral studies have found evidence that periods of inattention ('zoning out'), and not periods of split attention, are specifically associated with negative consequences for language comprehension as well as other cognitive tasks (Smallwood et al., 2008; Smallwood, McSpadden, et al., 2007). Overall, the available data suggest that the impact of periods of split attention on cognitive processing may be complex, but that split attention can reduce a reader or listener's sensitivity to context during language comprehension, as in the current study and in Hubbard and Federmeier (2021).

Why might this be the case? It is possible that participants were simply not very good at self-monitoring their own attentional state, and may have used the response options corresponding to inattentive and split attention states somewhat interchangeably. While this cannot completely be ruled out, there is some data that suggests otherwise. Previous analyses of a subset of the current dataset found a significant difference in neural oscillations associated with attention when comparing inattentive to split attention states (Boudewyn & Carter, 2018). Specifically, neural oscillations in the alpha frequency band[1] were increased when participants were inattentive (relative to when they were attentive), but were decreased when they were in a state of split attention. In addition, previous work from other groups has found behavioral differences on language comprehension and executive control tasks between inattentive and split attention states (Smallwood et al., 2008; Smallwood, McSpadden, et al., 2007). This suggests that participants were successfully distinguishing between the different attentional states.

Thus, in keeping with a large and growing literature showing the costs of dividing attention, these results may indicate that 'half-listening' and dividing attention during language comprehension has costs on par with fully 'zoning out' and losing attentional focus. This is consistent with the suggestion that truly split attention is an illusion, and that when we attempt to split our attention or multi-task, the brain is instead rapidly toggling between tasks (Miller & Buschman, 2015). Given the dependence of language comprehension on maintaining attentional focus over time, this kind of shifting between the task and another locus of attention would be expected to impact the ability to construct a representation of the overall context and to make use of that representation to processing incoming words. This fits with the pattern of results observed in this study, in which 'split attention' leads to the same pattern of word processing as 'inattention' in terms of sensitivity to relatively global context, but with intact sensitivity to local word level context.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

Armeni, K., Willems, R. M., & Frank, S. L. (2017). Probabilistic language models in cognitive neuroscience: Promises and pitfalls. *Neuroscience & Biobehavioral Reviews*, *83*, 579–588. https://doi.org/10.1016/j.neubiorev.2017.09.001

Boudewyn, M. A., & Carter, C. S. (2018). I must have missed that: Alpha-band oscillations track attention to spoken language. *Neuropsychologia*, *117*, 148–155. https://doi.org/10.1016/j.neuropsychologia.2018.05.024

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2012). Cognitive control influences the use of meaning relations during spoken sentence comprehension. *Neuropsychologia*, *50*(11), 2659–2668. https://doi.org/10.1016/j.neuropsychologia.2012.07.019

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2013). Effects of working memory span on processing of lexical associations and congruence in spoken discourse. *Frontiers in Psychology*, *4*, 60. https://doi.org/10.3389/fpsyg.2013.00060

Boudewyn, M. A., Long, D. L., Traxler, M. J., Lesh, T. A., Dave, S., Mangun, G. R., Carter, C. S., & Swaab, T. Y. (2015). Sensitivity to referential ambiguity in discourse: The role of attention, working memory, and verbal ability. *Journal of Cognitive Neuroscience*, *27*(12), 2309–2323. https://doi.org/10.1162/jocn_a_00837

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Cristescu, T. C., & Nobre, A. C. (2008). Differential modulation of word recognition by semantic and spatial orienting of attention. *Journal of Cognitive Neuroscience*, *20*(5), 787–801. https://doi.org/10.1162/jocn.2008.20503

Doyle, A. C. (1930). *The complete Sherlock Holmes*. Doubleday & Co.

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*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

[1]See Boudewyn and Carter (2018) for a more comprehensive discussion of the functional significance of alpha oscillations. In brief: alpha oscillations tend to be reduced when attention is directed to an external stimulus or task, and tend to increase when disengaged from external stimuli.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006
Frank, S. L., & Willems, R. M. (2017). Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension. *Language, Cognition and Neuroscience*, *32*(9), 1192–1203. https://doi.org/10.1080/23273798.2017.1323109
Franklin, M. S., Broadway, J. M., Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2013). *Window to the wandering mind: Pupillometry of spontaneous thought while reading*. SAGE Publications Sage UK.
He, T., Boudewyn, M. A., Kiat, J. E., Sagae, K., & Luck, S. J. (2022). Neural correlates of word representation vectors in natural language processing models: Evidence from representational similarity analysis of event-related brain potentials. *Psychophysiology*, *59*(3), e13976. https://doi.org/10.1111/psyp.13976
Hubbard, R. J., & Federmeier, K. D. (2021). Dividing attention influences contextual facilitation and revision during language comprehension. *Brain Research*, *1764*, 147466. https://doi.org/10.1016/j.brainres.2021.147466
Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299
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*(1), 621. https://doi.org/10.1146/annurev.psych.093008.131123
Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213. https://doi.org/10.3389/fnhum.2014.00213
Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, *383*(6601), 616–618. https://doi.org/10.1038/383616a0
McCarthy, G., & Nobre, A. C. (1993). Modulation of semantic processing by spatial selective attention. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *88*(3), 210–219. https://doi.org/10.1016/0168-5597(93)90005-A
Michaelov, J. A., Bardolph, M. D., Coulson, S., & Bergen, B. K. (2021). Different kinds of cognitive plausibility: Why are transformers better than RNNs at predicting N400 amplitude? In Proceedings of the Annual Meeting of the Cognitive Science Society. 43. https://doi.org/10.48550/arXiv.2107.09648.
Michaelov, J. A., & Bergen, B. K. (2022). Do language models make human-like predictions about the coreferents of Italian anaphoric zero pronouns? (arXiv: 2208.14554). *arXiv*. https://doi.org/10.48550/arXiv.2208.14554
Miller, E. K., & Buschman, T. J. (2015). Working memory capacity: Limits on the bandwidth of cognition. *Daedalus*, *144*(1), 112–122. https://doi.org/10.1162/DAED_a_00320
Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAi Blog*, *1*(8), 9. https://github.com/openai/gpt-2
Ruz, M., & Nobre, A. C. (2008). Attention modulates initial stages of visual word processing. *Journal of Cognitive Neuroscience*, *20*(9), 1727–1736. https://doi.org/10.1162/jocn.2008.20119
Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences*, *15*(7), 319–326. https://doi.org/10.1016/j.tics.2011.05.006
Smallwood, J. (2011). Mind-wandering while reading: Attentional decoupling, mindless reading and the cascade model of inattention. *Language and Linguistics Compass*, *5*(2), 63–77. https://doi.org/10.1111/j.1749-818X.2010.00263.x
Smallwood, J., Fishman, D. J., & Schooler, J. W. (2007). Counting the cost of an absent mind: Mind wandering as an underrecognized influence on educational performance. *Psychonomic Bulletin & Review*, *14*(2), 230–236. https://doi.org/10.3758/BF03194057
Smallwood, J., McSpadden, M., & Schooler, J. W. (2007). The lights are on but no one's home: Meta-awareness and the decoupling of attention when the mind wanders. *Psychonomic Bulletin & Review*, *14*(3), 527–533. https://doi.org/10.3758/BF03194102
Smallwood, J., McSpadden, M., & Schooler, J. W. (2008). When attention matters: The curious incident of the wandering mind. *Memory & Cognition*, *36*(6), 1144–1150. https://doi.org/10.3758/MC.36.6.1144
Szewczyk, J. M., & Federmeier, K. D. (2022). Context-based facilitation of semantic access follows both logarithmic and linear functions of stimulus probability. *Journal of Memory and Language*, *123*, 104311. https://doi.org/10.1016/j.jml.2021.104311