Mental effort during mindless reading? Pupil fluctuations indicate internal processing during levels of inattention

Daniel J. Schad, Antje Nuthmann, Frank Rösler, & Ralf Engbert

Supplementary Material

Type of Error	Example sentence
Control	**The class representative** (1) On all birthdays, he congratulates his classmates and the teacher.
Gibberish text	(1) On all classmates, he congratulates his birthdays and the teacher.
Discourse errors	(1) He sais hello to guests in the name of the class. He congratulates his classmates on their birthdays but never sais hello to guests.
Semantic errors	(1) He always thinks of buying new hamsters for the bathroom at school.
Control	**At the old city wall** (2) The wall was made from big worked stones.
Syntactic errors	(2) The wall worked was made from big stones.
Lexical errors	(2) The wall was begrothed from big worked stones.

Supplementary Figure S1: Example sentences for the different error types.

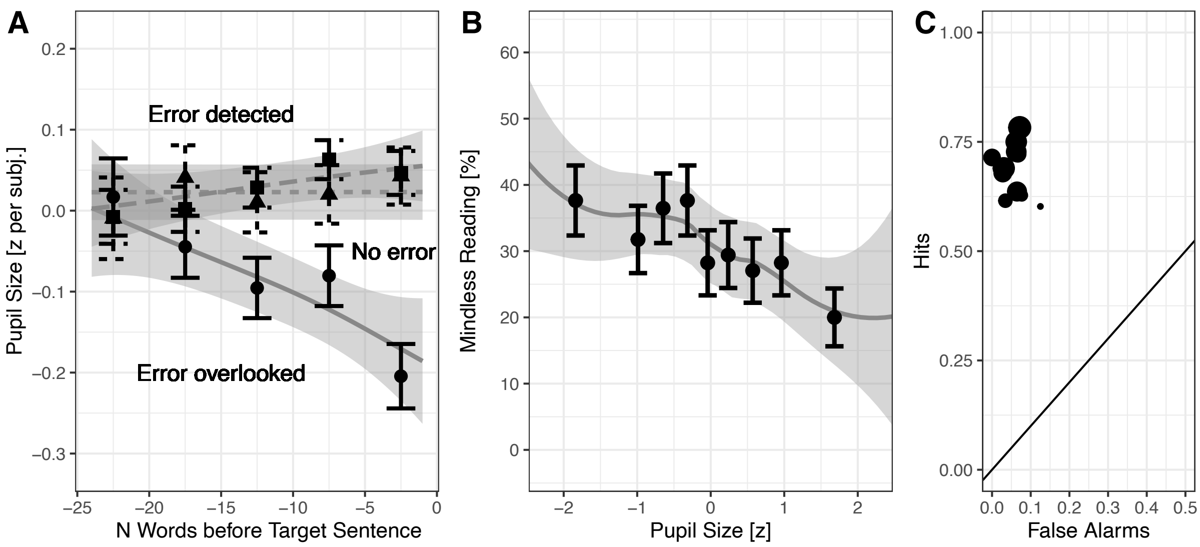
Supplementary Table S1: Control analyses for effects of error detection on pupil size (z-transformation across subjects).

	Pupil_size_z		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.17	-0.50 – 0.16	0.323
Error detected vs overlooked	0.14	0.05 – 0.23	**0.002**
No error vs error overlooked	0.10	0.00 – 0.19	**0.039**
N word prior to error sentence	-0.00	-0.01 – -0.00	**0.041**
Page number [1st degree]	-38.71	-41.15 – -36.27	**<0.001**
Page number [2nd degree]	9.92	9.02 – 10.82	**<0.001**
Page number [3rd degree]	-2.35	-3.58 – -1.11	**<0.001**
Page number [4th degree]	4.00	3.10 – 4.90	**<0.001**
Page number [5th degree]	-4.96	-6.00 – -3.92	**<0.001**
Page number [6th degree]	-1.44	-2.33 – -0.55	**0.001**
Line number [1st degree]	-4.94	-10.52 – 0.64	0.083
Line number [2nd degree]	-6.23	-7.51 – -4.94	**<0.001**
Line number [3rd degree]	6.81	5.58 – 8.03	**<0.001**
Line number [4th degree]	-1.93	-2.94 – -0.91	**<0.001**
Line number [5th degree]	-0.19	-1.16 – 0.78	0.698
Line number [6th degree]	1.07	0.16 – 1.98	**0.021**
Line number [7th degree]	-0.59	-1.50 – 0.32	0.206
Word number per line [1st degree]	-0.35	-1.48 – 0.79	0.549
Word number per line [2nd degree]	-2.95	-3.82 – -2.07	**<0.001**
Word number per line [3rd degree]	0.75	-0.12 – 1.62	0.090
Word number per line [4th degree]	0.88	0.02 – 1.73	**0.045**
Word number per line [5th degree]	-1.77	-2.62 – -0.92	**<0.001**
Fixation duration [degree]	1.27	0.38 – 2.16	**0.005**
Saccade length [1st degree]	-1.29	-2.17 – -0.40	**0.004**
Saccade length [2nd degree]	-1.22	-2.14 – -0.29	**0.010**
Saccade length [3rd degree]	-0.95	-1.80 – -0.09	**0.031**
Saccade length [4th degree]	-0.79	-1.66 – 0.08	0.076
Word frequency [degree]	-0.93	-2.30 – 0.43	0.180
Word length [1st degree]	-0.44	-1.85 – 0.98	0.547
Word length [2nd degree]	0.93	-0.00 – 1.86	0.051
Error detected vs overlooked × N word prior to error sentence	0.00	0.00 – 0.01	**0.006**
No error vs error overlooked × N word prior to error sentence	0.00	-0.00 – 0.01	0.177
Random Effects			
σ^2	0.18		
τ_{00} Text_page	0.05		
τ_{00} Subject.5	0.74		
τ_{11} Subject.No_error_vs_error_overlooked:N_word_prior_to_error_sentence	0.00		
τ_{11} Subject.Error_detected_vs_overlooked:N_word_prior_to_error_sentence	0.00		
τ_{11} Subject.N_word_prior_to_error_sentence	0.00		
τ_{11} Subject.No_error_vs_error_overlooked	0.04		
τ_{11} Subject.Error_detected_vs_overlooked	0.04		
ρ_{01}			
ρ_{01}			
ICC	0.24		
N Subject	27		
N Text_page	62		
Observations	13290		
Marginal R^2 / Conditional R^2	0.340 / 0.496		

Table note. Polynomial effects of covariates were computed (and scaled) using the "poly" function in R.

Supplementary Material: Analyses with z-transformation of pupil size for subject separately (i.e., pure within-subject analyses)

We next performed the z-transformation for each subject individually. Thus, all analyses of pupil size purely capture within-subject variation in pupil size. The overall pattern of results remained unchanged; i.e., the results were stable with respect to the precise definition of the pupil size variable.



Supplementary Figure S2. Replicating parts of Figure 3, but performing the z-transformation of pupil size for each subject separately rather than across subjects. (A) Pupil size before errors were overlooked (i.e., during mindless reading, solid bottom line), before errors were detected (i.e., mindful reading, dashed upper line), and in a control condition (no error, dotted middle line) for different number of words prior to the error sentence. (B) Percent overlooked errors (i.e., mindless reading) as a function of average pupil size on the five words prior to the target sentence. For each subject, average pupil size was computed for each page of text containing an error; text pages where the target sentence contained no error were excluded. The average pupil size per text page was divided into 9 equally spaced bins across subjects based on quantile splits. Average pupil size for each bin is plotted on the x-axis, and the associated percentage of overlooked errors in shown on the y-axis. (A+B) Local regression models (LOESS) and associated standard errors for the ungrouped data are displayed, as well as means and standard error of the mean (SEM) for bins, each containing data from at least 400 fixations (panel A) or 85 text pages (panel B). (C) Hits (i.e., correctly detected errors) over false alarms (i.e., falsely detected errors) for different pupil sizes, where large points indicate bins with larger pupil sizes.

First, we tested effects of error detection averaged across all error types for each experiment separately. We found that in experiment 1, when errors where overlooked (versus detected), pupil size at the onset of the error sentence was significantly reduced (b = 0.271, SE = 0.098, t = 2.76, p = .009), and this effect got stronger with more proximity to the error sentence (b = 0.015, SE = 0.005, t = 2.92, p = .006). Similar findings were present in experiment 2. Pupil size at the onset of the error sentence was reduced when errors were overlooked (b = 0.336, SE = 0.101, t = 3.32, p = .002). However, the increase of this effect when getting closer to the error sentence was only marginally significant (b = 0.011, SE = 0.005, t = 2.00, p = .051).

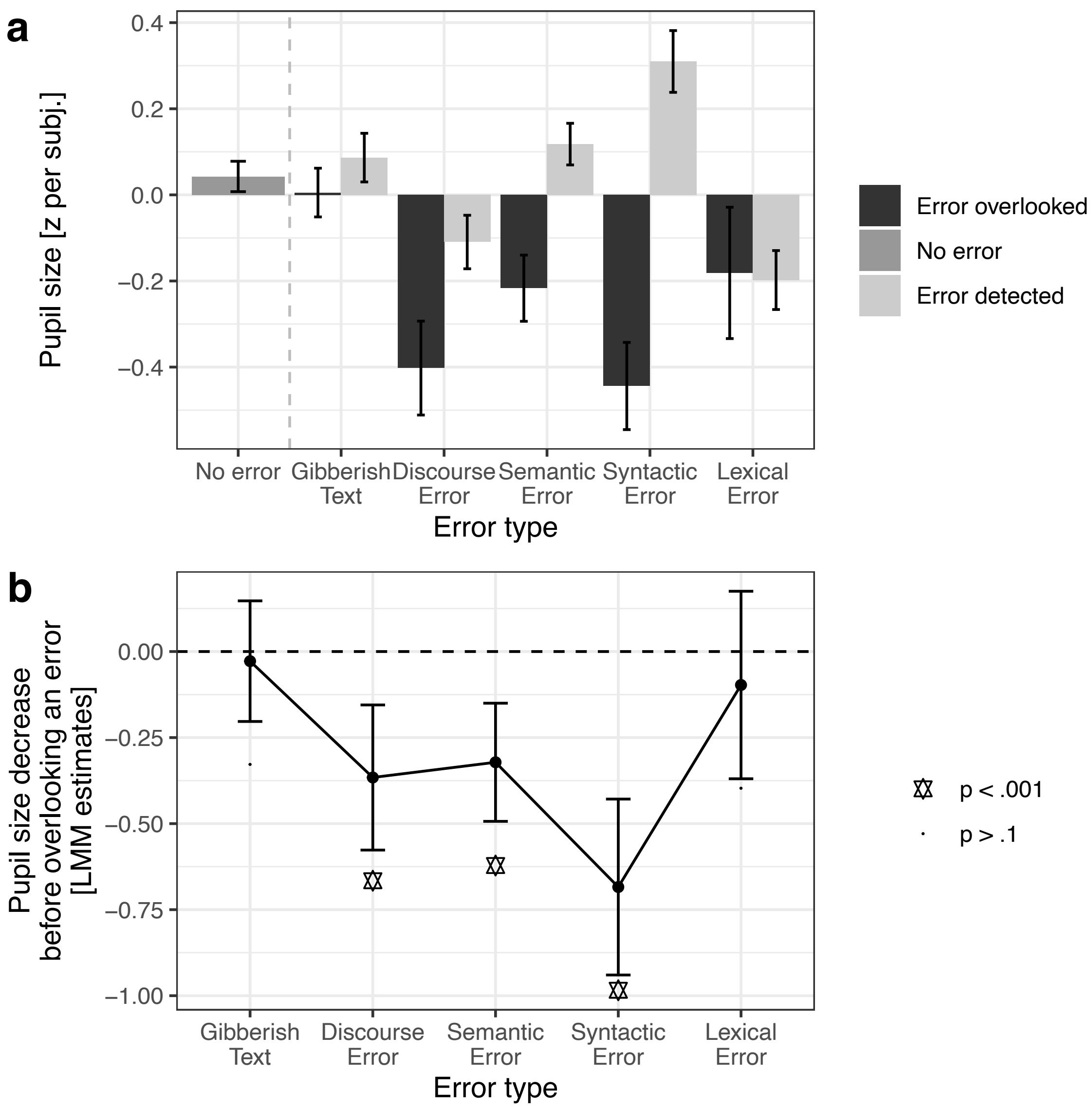
Averaged over both experiments (see Fig. S2), pupils at the onset of the error sentence were again significantly smaller when errors were overlooked than when they were detected (simple effect for “error detection”: b = 0.300, SE = 0.077, t = 3.90, p < .001; comparison of “overlooked” to the control condition with “no error” was now significant – contrary to the primary analysis reported in the manuscript text: b = 0.231, SE = 0.109, t = 2.11, p = .047). As in the primary analysis, the effect of overlooking errors became stronger as readers approached the error sentence (i.e., interaction “error overlooked” vs. “error detected” × word number: b = 0.011, SE = 0.004, t = 2.60, p = .014), indicating that pupil size was temporally closely associated with zoning out. We again conducted control analyses to ensure that these effects were not driven by certain aspects of the task, design, or word characteristics. The results substantiated that the effects described above were highly robust when controlling for higher-order polynomials of time on task, line number, word number, fixation duration, saccade amplitude, word length, and word frequency (see Table S2).

Supplementary Table S2: Control analyses for effects of error detection on pupil size (z-transformation per subject).

	Pupil_size_z_per_subject		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.24	-0.39 – -0.09	**0.002**
Error detected vs overlooked	0.25	0.11 – 0.39	**<0.001**
No error vs error overlooked	0.19	0.05 – 0.33	**0.009**
N word prior to error sentence	-0.01	-0.01 – 0.00	0.059
Page number [1st degree]	-75.66	-80.55 – -70.77	**<0.001**
Page number [2nd degree]	21.55	19.74 – 23.37	**<0.001**
Page number [3rd degree]	-5.17	-7.66 – -2.68	**<0.001**
Page number [4th degree]	9.01	7.19 – 10.83	**<0.001**
Page number [5th degree]	-9.78	-11.88 – -7.69	**<0.001**
Page number [6th degree]	-1.50	-3.30 – 0.30	0.101
Line number [1st degree]	-11.10	-21.78 – -0.41	**0.042**
Line number [2nd degree]	-15.30	-17.89 – -12.70	**<0.001**
Line number [3rd degree]	14.33	11.85 – 16.81	**<0.001**
Line number [4th degree]	-4.47	-6.52 – -2.42	**<0.001**
Line number [5th degree]	-0.45	-2.41 – 1.51	0.655
Line number [6th degree]	2.11	0.27 – 3.95	**0.024**
Line number [7th degree]	-1.48	-3.32 – 0.36	0.115
Word number per line [1st degree]	-0.64	-2.90 – 1.62	0.578
Word number per line [2nd degree]	-6.73	-8.51 – -4.95	**<0.001**
Word number per line [3rd degree]	1.95	0.19 – 3.70	**0.030**
Word number per line [4th degree]	2.26	0.53 – 4.00	**0.011**
Word number per line [5th degree]	-3.82	-5.55 – -2.09	**<0.001**
Fixation duration [degree]	2.84	1.04 – 4.64	**0.002**
Saccade length [1st degree]	-2.63	-4.42 – -0.83	**0.004**
Saccade length [2nd degree]	-1.83	-3.70 – 0.04	0.055
Saccade length [3rd degree]	-1.81	-3.55 – -0.07	**0.041**
Saccade length [4th degree]	-1.54	-3.32 – 0.23	0.088
Word frequency [degree]	-2.24	-5.01 – 0.53	0.113
Word length [1st degree]	-1.00	-3.87 – 1.87	0.494
Word length [2nd degree]	1.73	-0.15 – 3.62	0.071
Error detected vs overlooked × N word prior to error sentence	0.01	0.00 – 0.02	**0.006**
No error vs error overlooked × N word prior to error sentence	0.01	0.00 – 0.01	**0.047**
Random Effects			
σ^2	0.75		
$\tau_{00\ \mathrm{Text\_page}}$	0.19		
$\tau_{00\ \mathrm{Subject.5}}$	0.04		
$\tau_{11\ \mathrm{Subject.No\_error\_vs\_error\_overlooked:N\_word\_prior\_to\_error\_sentence}}$	0.00		
$\tau_{11\ \mathrm{Subject.Error\_detected\_vs\_overlooked:N\_word\_prior\_to\_error\_sentence}}$	0.00		
$\tau_{11\ \mathrm{Subject.N\_word\_prior\_to\_error\_sentence}}$	0.00		
$\tau_{11\ \mathrm{Subject.No\_error\_vs\_error\_overlooked}}$	0.08		
$\tau_{11\ \mathrm{Subject.Error\_detected\_vs\_overlooked}}$	0.09		
ρ_{01}			
ρ_{01}			
ICC	0.20		
N_{Subject}	27		
$N_{\mathrm{Text\_page}}$	62		
Observations	13290		
Marginal R^2 / Conditional R^2	0.338 / 0.473		

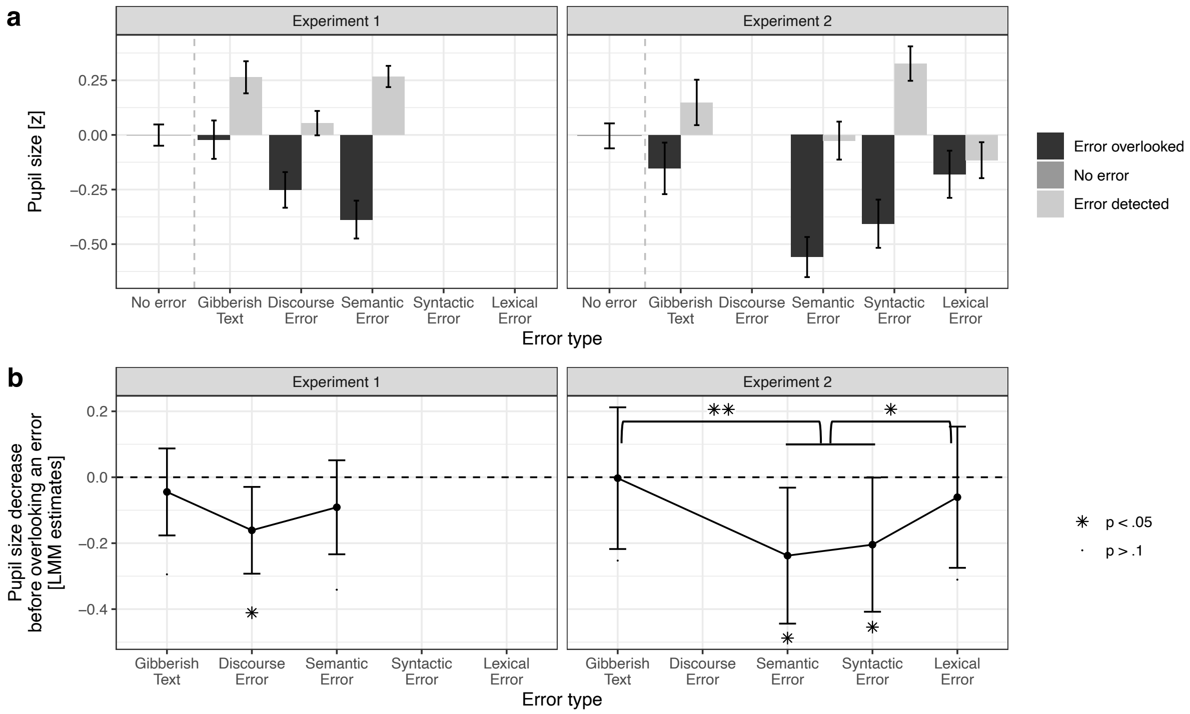
We also tested whether average pupil size on the five words prior to an error sentence can be used to predict overlooking the upcoming error. As is displayed in Figure S1b, small pupils predicted overlooking of an error, whereas large pupils predicted detection of an error (binomial GLME: b = -0.23, SE = 0.08, z = -2.83, p = .005).

A mixed effects signal detection analysis (Wright, Horry, & Skagerberg, 2009) revealed that pupil size in a region ten words prior to the target sentence predicted sensitivity for errors marginally ($\Delta d'$ = 0.19, SE = 0.13, z = 1.45, $p_{1\text{-tailed}}$ = .07), but no response bias (Δc = 0.03, SE = 0.07, z = 0.38, p = .704). These results suggest that – as in the primary analysis – smaller pupils may be associated with reduced sensitivity for errors, i.e., with mindless reading. However, in contrast to the primary analysis, this effect was only marginally significant.



Supplementary Figure S3. Analysis of pupil size (z-standardized per subject) before different types of errors are overlooked. Overlooking different types of errors assesses different levels of attentional decoupling, ranging from weak decoupling (gibberish text and discourse errors) over medium decoupling (semantic and syntactic errors) to deep decoupling (lexical errors). A control condition included no error. Pupil size was measured on 5 words prior to the error sentence. (A) Average pupil size when an upcoming error was overlooked, was detected, or when no error was presented. Error bars are standard errors of the mean (SEM). (B) Decrease of pupil size before an error was overlooked as compared to when it was detected, as estimated by a linear mixed-effects model. Error bars are 95% confidence intervals. Results show a steadily stronger decrease of pupil size for more low-level errors (i.e., deeper levels of decoupling), but that pupil size was not reduced at the lowest level, when lexical errors are overlooked (deep decoupling), consistent with internal resource-demanding cognitive processing during deep decoupling.

The critical question was whether pupil size was reduced before different types of errors were overlooked as compared to when they were detected (see Fig. S3). We found that pupil size was not significantly reduced when high-level gibberish text was overlooked (b = -0.028, SE = 0.089, t = -0.31, p = .754). However, pupil size was significantly reduced when discourse errors were overlooked (b = -0.366, SE = 0.107, t = -3.41, p = .0007), when semantic errors were overlooked (b = -0.322, SE = 0.087, t = -3.68, p = .0002), and when syntactic errors were overlooked (b = -0.684, SE = 0.130, t = -5.25, p =.0000001). We then used custom simple difference contrasts, where one contrast compared medium-level errors (semantic and syntactic) to high-level errors (gibberish text and context errors), and the other contrast compared low-level (lexical) errors to medium-level errors (semantic and syntactic). The decrease of pupil size was marginally stronger for medium-level errors (semantic and syntactic) compared to high-level errors (gibberish text and semantic errors; for the difference: b = -0.315, SE = 0.156, t = -2.02, p = .053; this comparison was significant in the primary analysis). Critically, pupil size was not reduced when low-level lexical errors were overlooked (b = -0.097, SE = 0.139, t = -0.70, p = .484), which – as in the primary analysis – significantly differed from medium-level (semantic and syntactic) errors (b = 0.436, SE = 0.187, t = 2.33, p = .021). These results are in line with the decoupling hypothesis.



Supplementary Figure S4. Analysis of pupil size (z-standardized across subjects) before different types of errors are overlooked; split up by experiment. Overlooking different types of errors assesses different levels of attentional decoupling, ranging from weak decoupling (gibberish text and discourse errors) over medium decoupling (semantic and syntactic errors) to deep decoupling (lexical errors). A control condition included no error. Pupil size was measured on 5 words prior to the error sentence. (A) Average pupil size when an upcoming error was overlooked, was detected, or when no error was presented. Error bars are standard errors of the mean (SEM). (B) Decrease of pupil size before an error was overlooked as compared to when it was detected, as estimated by a linear mixed-effects model (LMM). Error bars are 95% confidence intervals. Results for Experiment 1 were stable for discourse errors but not for semantic errors; results for experiment 2 reflect the overall results: they show a stronger decrease of pupil size for more low-level errors (i.e., deeper levels of decoupling), but that pupil size is not reduced at the lowest level, when lexical errors are overlooked (deep decoupling), consistent with internal resource-demanding task-unrelated thought during deep decoupling.

Word-frequency effects as a marker of lexical processing difficulty

Research on eye-movement control in reading has shown that fixation durations decrease with increasing word frequency of the currently fixated word n (immediacy effect) (e.g., Inhoff & Rayner, 1986; Kliegl, Grabner, Rolfs, & Engbert, 2004). Corpus analyses additionally revealed evidence for distributed processing such that fixation times on word n were also modulated by the frequency of the neighboring words (i.e., word n - 1 and word n + 1) (Kliegl, Nuthmann, & Engbert, 2006). A word-frequency immediacy effect has also been reported for pupil size, with the pupil dilating more strongly in response to difficult (i.e., low-frequency) as compared to easy (i.e., high-frequency) words (Fernández, Biondi, Castro, & Agamenonni, 2016).

Here, we analyzed the pupil-size data for 10 words prior to a lexical error. Three separate LMMs assessed the frequency effects of word n, word n – 1, and word n + 1 on pupil dilation, nested under the mindful reading condition. For these analyses, we conducted one-tailed tests: we hypothesized that lexical influences on pupil size should be larger when lexical errors were detected compared to when they were not detected (Fernández et al., 2016; Kliegl et al., 2006; Smallwood et al., 2012; Smallwood et al., 2011).

The results from the first LMM indicate that pupil size was not modulated by word n frequency during mindful reading, that is before a lexical error was detected (b = 0.003, SE = 0.011, t = 0.31, p = .760). Given the slowness of the pupillary response (Loewenfeld, 1958), the difficulty of word n - 1 may spill over into pupil size on word n. The results from the second

LMM suggest, however, that word n - 1 frequency did not influence pupil dilation (*b* = -0.001, *SE* = 0.010, *t* = -0.72, *p* = .471). In the third LMM, we tested the influence of word n + 1 frequency, again nested under mindful reading. Notably, pupil size was larger for low-frequency than for high-frequency words n + 1 (*b* = -0.022, *SE* = 0.011, *t* = -2.04, *p* = .042). Importantly, this effect was absent before lexical errors were overlooked (nested effect; *b* = 0.018, *SE* = 0.019, *t* = 0.94, *p* = .348; for the difference between mindful vs. mindless reading, tested as the interaction of word n + 1 frequency × mindful vs. mindless: *b* = 0.040, *SE* = 0.021, *t* = 1.86, $p_{\text{1-tailed}}$ = .032). These results lend further support to the view that relatively large pupils during deep decoupling do not reflect external processing (TRT; i.e., phasic bursts), but may reflect internal cognitive activity (TUT).

Moreover, we investigated word-frequency effects using z-transformations of pupil size for each individual subject. We again tested the influence of word n + 1 frequency, again nested under mindful reading. Again, pupil size was larger for low-frequency than for high-frequency words n + 1 (*b* = -0.049, *SE* = 0.022, *t* = -2.23, *p* = .026). Importantly, this effect was absent before lexical errors were overlooked (nested effect; *b* = 0.059, *SE* = 0.039, *t* = 1.53, *p* = .126; for the difference between mindful vs. mindless reading, tested as the interaction of word n + 1 frequency × mindful vs. mindless: *b* = 0.108, *SE* = 0.044, *t* = 2.48, $p_{\text{1-tailed}}$ = .007). These results lend further support to the view that relatively large pupils during deep decoupling do not reflect external processing (TRT; i.e., phasic bursts), but may reflect internal cognitive activity (TUT). Overall, given that frequency effects occurred only for word n + 1, but not for words n and n – 1, frequency effects await further replication.

Supplementary References

Fernández, G., Biondi, J., Castro, S., & Agamenonni, O. (2016). Pupil size behavior during online processing of sentences. *Journal of Integrative Neuroscience, 15*(04), 485-496. doi:10.1142/S0219635216500266

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40*(6), 431-439. doi:10.3758/BF03208203

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16*(1-2), 262-284. doi:10.1080/09541440340000213

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology-General, 135*(1), 12-35. doi:10.1037/0096-3445.135.1.12

Loewenfeld, I. E. (1958). Mechanisms of reflex dilatation of the pupil. *Documenta Ophthalmologica, 12*(1), 185-448. doi:10.1007/bf00913471

Smallwood, J., Brown, K. S., Baird, B., Mrazek, M. D., Franklin, M. S., & Schooler, J. W. (2012). Insulation for daydreams: a role for tonic norepinephrine in the facilitation of internally guided thought. *PLoS One, 7*(4), e33706. doi:10.1371/journal.pone.0033706

Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., . . . Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PLoS One, 6*(3). doi:10.1371/journal.pone.0018298

Wright, D. B., Horry, R., & Skagerberg, E. M. (2009). Functions for traditional and multilevel approaches to signal detection theory. *Behavior Research Methods, 41*(2), 257-267. doi:10.3758/Brm.41.2.257