



Neural signatures for active maintenance and interference during working memory updating

Adrià Vilà-Balló^{a,b,c,d,*,1}, Juha Salmi^{a,1}, Anna Soveri^a, Antoni Rodríguez-Fornells^{b,e,f}, Minna Lehtonen^a, Matti Laine^{a,g}

^a Department of Psychology, Åbo Akademi University, Tehtaankatu 2, Turku 20500, Finland

^b Cognition and Brain Plasticity Group [Bellvitge Biomedical Research Institute-] IDIBELL, L'Hospitalet de Llobregat, Barcelona 08097, Spain

^c Department of Psychology, Faculty of Education and Psychology, University of Girona, Girona 17071, Spain

^d Headache and Neurological Pain Research Group, Vall d'Hebron Research Institute (VHIR), Barcelona 08035, Spain

^e Department of Cognition, Development and Educational Psychology, Faculty of Psychology, University of Barcelona, Barcelona 08035, Spain

^f Catalan Institution for Research and Advanced Studies, ICREA, Barcelona, Spain

^g Brain and Mind Center, University of Turku, Finland

ARTICLE INFO

Keywords:

EEG
Working memory
Updating
Interference
Positive slow wave
N-back task

ABSTRACT

Although working memory (WM) is amongst the most studied neurocognitive functions, temporal patterns of its component processes are not fully understood. We examined the neural underpinnings of active maintenance and interference management in the n-back task by manipulating load (1-back vs 3-back) and including so-called lure stimuli. ERPs of 27 young adults revealed that the 1-back condition enabling active maintenance showed a positive slow wave (PSW) prior to the next stimulus (−600–0 ms) and augmented P2 (190–290 ms) and P3b (330–430 ms) responses after the stimulus appeared, albeit the latter effects were driven by the initial PSW. Moreover, PSW amplitude correlated negatively with reaction time in the 1-back condition. Responses to lures showed interference, accompanied with different ERP effects for the two load levels. Our results support the view that PSW reflects efficient WM maintenance and suggest two distinct neuronal correlates for interference in WM.

1. Introduction

Flexible updating of working memory (WM) contents is considered one of the key aspects of executive functioning (Miyake et al., 2000). Updating entails refreshment of WM contents so that the information maintained is relevant to the task or goal at hand, as well as selection of incoming information, inhibition of the irrelevant information, and continuous monitoring of performance (e.g., Morris & Jones, 1990). Thus, successful WM updating requires both maintenance and executive operations. The role of maintenance vs. executive operations in WM updating is thought to vary depending on the task demands: when the executive demands are low, there are more resources available for the active maintenance of relevant information, and vice versa (Ecker, Lewandowsky, Oberauer, & Chee, 2010; Ecker, Maybery, & Zimmer, 2013; Botto, Basso, Ferrari, & Palladino, 2014). In the present study, we employed the n-back task commonly used in WM research and simultaneously registered event-related brain potentials (ERPs) under varying task load. We focused on the active maintenance component

(pre-stimulus period) and management of interference, aiming to examine the neurophysiological correlates underlying these two key WM functions.

Originally developed by Kirchner in 1958, the n-back is a continuous performance task that requires the participant to decide whether the current stimulus matches the one *n* steps back in the stimulus sequence. Simply by changing the instruction, the researcher can parametrically vary the WM load from a minimal (1-back) load upwards. Stimulus sequences can also contain lure stimuli that would be targets for adjacent load levels. Lures tap the management of interference in WM as they prolong reaction times and decrease accuracy (e.g., Kane, Conway, Miura, & Colflesh, 2007; Schmiedek, Li, & Lindenberger, 2009). In addition to parametric design and strict control of stimulation measures, the advantages of the n-back tasks include high concurrent validity (Jaeggi, Buschkuhl, Perrig, & Meier, 2010), and high reliability reported especially in brain imaging studies (Owen, McMillan, Laird, & Bullmore, 2005). Although n-back does not correlate strongly with span tasks and both measures appear to account for

* Corresponding author at: Cognition and Brain Plasticity Unit (Campus Bellvitge), Department of Cognition, Development and Education Psychology, Faculty of Psychology, University of Barcelona, Barcelona 08035, Spain.

E-mail address: adria.vila@vhir.org (A. Vilà-Balló).

¹ Authors contributed equally to the manuscript.

independent variance in fluid intelligence (Kane et al., 2007), evidence from latent factor analyses has indicated that n-back tasks are valid indicators of general WM function (Schmiedek, Lövdén, & Lindenberger, 2014). This paradigm has been widely employed, for instance, in the study of the neuronal correlates of WM (Owen et al., 2005) and in training of WM (Au et al., 2015; Soveri, Antfolk, Karlsson, Salo, & Laine, 2017).

Considering the key role of updating in WM, it is surprising that only a handful of studies have examined ERP responses during the n-back tasks. In the 1-back condition, the participant should be able to keep the previous stimulus actively in mind as it is the one that will be matched with the immediately following stimulus. However, in a clearly more demanding condition, such as the 3-back, the decision on whether or not a trial is the same as three trials back likely calls for recall of the previous relevant stimulus, as its constant maintenance is getting difficult when there are intervening stimuli (see Postle, Druzgal, & D'Esposito, 2003; Bunting, Cowan, & Saults, 2006). The ERP findings in the 1-back condition, but less so in the 3-back condition, should therefore show evidence for constant active WM maintenance between the stimuli.

Based on previous research, including the popular delayed matching-to-sample paradigm probing short-term memory (e.g., Ruchkin, Canoune, & Ritter, 1995; Mecklinger and Müller, 1996; Mecklinger & Pfeifer, 1996), maintenance of information in mind could be reflected as a slow-wave component that appears in-between stimuli. Both positive and negative slow waves with varying topographies have been reported during short-term memory task performance and tentatively assigned to different functional roles (see Ruchkin, Grafman, Canoune, & Ritter, 1992; Ruchkin et al., 1995). Consistent with these findings, several studies in non-human primates have reported anticipatory activity during retention period of the delayed matching-to-sample task (see Fuster, 2000 for a review). Similarly, previous studies have observed slow negative components in posterior electrode locations persisting throughout the retention interval in visual working memory tasks (Klaver, Talsma, Wijers, Heinze, & Mulder, 1999), and that were sensitive to the number of objects to be maintained in memory (Fukuda, Awh, & Vogel, 2010; Luria, Balaban, Awh, & Vogel, 2016). Similarly, in a recent MEG/EEG study (Albouy, Weiss, Baillet, & Zatorre, 2017) that used an auditory task requiring the comparison of two different patterns in temporal order, the authors showed an increased negative slow wave component that was associated to manipulation of information during the retention period. This theta oscillatory component also correlated with participants' performance accuracy. Importantly for the present purposes, a posteriorly located positive slow wave (PSW) related to active maintenance of WM contents (Ruchkin, Canoune, & Ritter, 1990; Rösler & Heil, 1991) has been recently observed between n-back trials as well (Bailey, Mlynarczyk, & West, 2016). In this study, PSW was strongest for the 1-back condition during which the participants are expected to actively maintain the contents of the previous stimulus. Bailey and colleagues also showed that the PSW distinguished between 1-back and 3-back hits, while it did not differentiate 3-back hits and misses. Moreover, they found that the PSW was insensitive to stimulus set size in 2-back and 3-back conditions. Altogether, these findings support the assumption that the PSW is involved in maintenance of information especially in the 1-back condition.

In the canonical P1, N1, P2, N2, P3 response sequence following each stimulus, especially P2 and P3 responses are modulated by cognitive load. P2 response (a positive waveform peaking at 200 ms post stimulus) has been linked to efficient selection of information at the sensory cortices (see Crowley & Colrain 2004). Modulation of P2 response has been reported in several types of WM tasks, from span tasks (e.g., Lefebvre, Marchand, Eskes, & Connolly, 2005; Marchand, Lefebvre, & Connolly, 2006) to delayed matching-to-sample (Ruchkin et al., 1995; Mecklinger & Müller, 1996), and updating (e.g., McEvoy, Smith, & Gevins, 1998; Rämä et al., 2000; Lenartowicz, Escobedo-

Quiroz, & Cohen, 2010; Luu et al., 2014; Dong, Reder, Yao, Liu, & Chen, 2015).

Amplitude modulation in P3 component is one of the most frequent findings in ERP studies focusing on WM updating. This is the case especially for its latter subcomponent, the P3b, which is associated to WM updating, allocation of attentional resources, and/or amount of resources demanded in the current task (Donchin, Miller, & Farwell, 1986; Dien, Spencer, & Donchin, 2004; Lenartowicz et al., 2010; Daffner et al., 2011). The modulation of the P3 component in the n-back could reflect the necessary cross-talk between memory, decision-making and orientation/change detection mechanisms (Chapman & Bragdon, 1964; Finnigan, Humphreys, Dennis, & Geffen, 2002; Polich, 2007; Rugg & Curran, 2007). Gevins et al. (1996) showed that updating modulates orthogonal source derivations (so-called Laplacian derivations) in EEG epochs at 300–1000 ms as compared with a simple target detection task. These responses were enhanced first at the posterior prefrontal sites and then at the posterior parietal sites. This investigation, as well as several other studies (e.g., Gevins et al., 1996; McEvoy et al., 1998) have reported larger responses to matching vs. non-matching stimuli at P3 latencies.

Due to multiple overlapping sources and several psychological factors modulating P2 and P3 responses, the WM load effects on these components have not been easy to delineate. For example, findings concerning WM-related modulations of P2 and P3 components are quite controversial. Furthermore, some studies have reported enhanced ERP amplitudes by increased WM load (see Marchand et al., 2006; Evans & Federmeier, 2007; Lefebvre et al., 2005; Chen, Mitra, & Schlaghecken, 2008; Chen & Mitra, 2009) while others have found the opposite pattern (McEvoy et al., 1998; Rämä et al., 2000; Dong et al., 2015; Bailey et al., 2016). Augmented ERP responses would go together with an extensive body of functional magnetic resonance imaging literature consistently showing that neuronal activity is increased in widespread areas when WM load increases (Owen et al., 2005). On the other hand, smaller P3 amplitude with increasing WM load is in line with ERP results from other types of tasks, indicating a reduced P3 amplitude when attentional demands increase or more resources are needed for the current task (e.g., Kramer, Wickens, & Donchin, 1987; Kok 2001). In line with this view, a recent study utilizing an independent component analysis for ERP data suggested that the posterior and anterior areas of the attention networks show specific associations with performance during the n-back tasks (Luu et al., 2014). In addition to these responses, WM-related modulation of ERPs has been observed at 600–1200 ms latencies (e.g., Marchand et al., 2006; Evans & Federmeier, 2007; Lefebvre et al., 2005; Rämä et al., 2000; Bailey et al., 2016).

Continuous updating prevents strong binding to an ongoing stimulus stream, leading to increased susceptibility to proactive interference (Kane et al., 2007; Schmiedek et al., 2009; Szmalec, Verbruggen, Vandierendonck, & Kemps, 2011). In the n-back task, interference effects are typically examined by using lure stimuli that would be targets at adjacent load levels. For instance, in a 3-back task including 2-back targets ($n - 1$ lure) as well as 4-back targets ($n + 1$ lure), both types of lures are associated with increased RTs and decreased accuracy (Kane et al., 2007; Schmiedek et al., 2009; Szmalec et al., 2011). Lures have been shown to modulate brain activity in multiple brain areas, and related activity further shares variance in performance in WM spans and Gf tasks (Burgess, Gray, Conway, & Braver, 2011). Preliminary evidence from a WM training study (Oelhafen et al., 2013) and a study with ADHD participants (Stroux et al., 2016) suggest that the interference effects resulting from binding of attention to items that are reminiscent of target items may occur at about 200–300 ms post-stimulus. However, it is not quite clear how interference influences brain functions at varying WM loads.

Following the idea that an important difference between the low- and high-load conditions in the n-back task lies in the possibilities for active maintenance of current WM contents, we examined both the pre-

and post-stimulus interval ERPs. Our first goal was to verify the findings of a recent important study by Bailey et al. (2016), who demonstrated that the low-load 1-back condition would trigger a slow-wave pre-stimulus ERP response, reflecting the continuous active maintenance of the previous stimulus in WM. As there is currently no direct evidence that this response reflects efficient maintenance of information, our second goal was to demonstrate a link between PSW response and behavioral performance as reflected by decision latencies. We hypothesized that if PSW reflects maintenance of information in the updating task as suggested by Bailey et al. (2016), its amplitude should predict the decision latency of the subsequent stimulus in the 1-back task where the conditions for maintenance are favorable. Finally, our third goal was to examine how interference elicited by lure stimuli influences brain responses at low and high WM load. Based on behavioral studies (Kane et al., 2007; Schmiedek et al., 2009), we expected that lure-related interference effects would be observed at both load levels. However, as these two tasks presumably engage WM subfunctions to different degrees, their neural signatures of interference could also differ. Moreover, the 3-back task allows the comparison between $n - 1$ vs $n + 1$ lure effects where the latter lure has just moved beyond the window for currently relevant stimuli.

2. Materials and methods

2.1. Participants

The current sample included 27 right-handed Spanish university students. None of them reported any neurological or psychiatric disorders or a history of severe head injury (Table 1). All participants gave their informed consent. The present data stem from the pretest of a WM training study (the results from that intervention study will be reported elsewhere), and the participants were reimbursed with 100 euros after the completion of the whole study. The study was reviewed and accepted by the Clinical Research Ethics Committee of the Bellvitge University Hospital, University of Barcelona, Spain.

2.2. Procedure

We used a digit n-back task adapted to simultaneous measurement of ERPs (see Fig. 1). The stimuli were numbers from 1 to 9, presented one at a time in the middle of a computer screen at a viewing distance of 70 cm. Each trial began with a fixation point. After 450 ms, a number appeared on the screen for 1500 ms. The stimulus onset asynchrony was fixed to 1950 ms. When the currently presented number was the same as the previous number (1-back task) or the number presented three numbers before (3-back task), the participant was to press the ‘yes’ button (target trial). For other numbers, one was to press the ‘no’ button (non-target trial). The task consisted of eight 1-back sequences and sixteen 3-back sequences. Each sequence consisted of 48 trials, resulting in altogether 1152 trials. To avoid merely familiarity-based decisions (e.g., Kane et al., 2007), we included lure trials in each sequence. Specifically, a lure trial in the 1-back condition was an $n + 1$ item that matched with the number presented two numbers back (e.g., in the sequence 2-3-4-3, the last number ‘3’ is a $n + 1$ lure trial). In the 3-back sequences, there were two types of lures, $n + 1$ (numbers that matched with the items presented four numbers back) and $n - 1$ (numbers that

matched with items presented two numbers back). In order to have enough epochs for the ERP analyses, the number of trials was set to 128 for each type of trials in the 1-back condition (target trials, standard non-target trials, and lure non-target trials), and to 256 for each type of trials in the 3-back condition (target, standard non-target, and lure non-target trials with $128 n - 1$ lures and $128 n + 1$ lures).

The order of the sequences was randomized for each participant. At the beginning of each sequence, there was a brief 10-s instruction indicating if the next block was a 1-back or 3-back block. The participants could also use these instruction periods to blink. The participants received two training sequences of 21 trials each, one with 1-back trials and the other one with 3-back trials. The total duration of the task was ~45 min.

2.3. Behavioral data analyses

The dependent variables were accuracy measures (d-prime for overall accuracy, correct rejection rates for non-targets) as well as RTs for correct trials in the two n-back tasks. The d' values were calculated following Stanislaw and Todorov (1999). Regarding the statistical analyses, repeated measures ANOVAs were at first performed on the accuracy rates and RTs with load (1-back; 3-back) and trial type (target; non-target) as within-subject factors. For these initial analyses, the two types of non-target trials (standard and lure) were lumped together. The overall analyses were followed by repeated measures ANOVAs that addressed performance on non-targets in more detail. These analyses were run separately for 1-back and 3-back. For 1-back, non-targets included standard non-targets and $n + 1$ lures, and for 3-back standard non-targets, $n + 1$ lures, and $n - 1$ lures. The whole sample could be used for the behavioral analyses, as no participant met the criteria of being an extreme outlier in accuracy or RTs (performance more than three times the interquartile range below or above the 1st or 3rd quartile, respectively).

2.4. Electrophysiological recording

Electroencephalogram (EEG) was recorded continuously (digitized with a sampling rate of 250 Hz, bandpass 0.01–70 Hz) using SynAmp Neuroscan amplifiers from 29 tin electrodes mounted on an elastic cap and located at standard positions (FP1/2, F3/4, C3/4, P3/4, O1/2, F7/8, T3/4, T5/6, Fz, Cz, Pz, FC3/4, FT7/8, CP3/4, TP7/8, FCz, CPz), and the left and right mastoids. The EEG was referenced on-line to the right ocular canthus and offline to the mean activity at the two mastoid electrodes. Vertical eye movements were monitored by an electrode placed below the right eye. Electrode impedances were maintained below 5 k Ω .

2.5. Preprocessing of the EEG data

Event-related potentials were time-locked to the stimulus presentation (time-range from -1000 to 2000 ms). First, waveform averages were separately obtained for each WM load (1-back, and 3-back) and trial type (target, and non-target trials). In an effort to avoid the suppression of PSW, a long pre-stimulus baseline period (-1000 to 800 ms) was used for the PSW analyses. Regarding the post-stimulus components (P2, P3b, and late negativity), a baseline from -100 to 0 ms was used. After this, to delineate the specific effects of each type of non-target trial in each condition, additional ERPs were obtained separately for each non-target type (1-back: standard non-target, and $n + 1$ lure; 3-back: standard non-target, $n + 1$ lure, and $n - 1$ lure). Importantly, only correct trials were used for the analyses.

Epochs exceeding ± 75 μ V in electrooculogram (EOG) or EEG were removed offline for further analysis using the extreme value function of the EEGlab toolbox. In the behavioral and ERP analyses, only the RT responses slower than 120 ms or faster than 3 standard deviations from the participant’s mean were considered for the analyses. Only artifact-

Table 1
Summary of Demographics and Average Scores on the Beck Depression Inventory-II (BDI-II).

N	27
Sex F/M	23/4
Age in years (SD)	21.48 (2.74)
Education in years (SD)	15.48 (1.53)
BDI-II (SD)	3.85 (3.53)

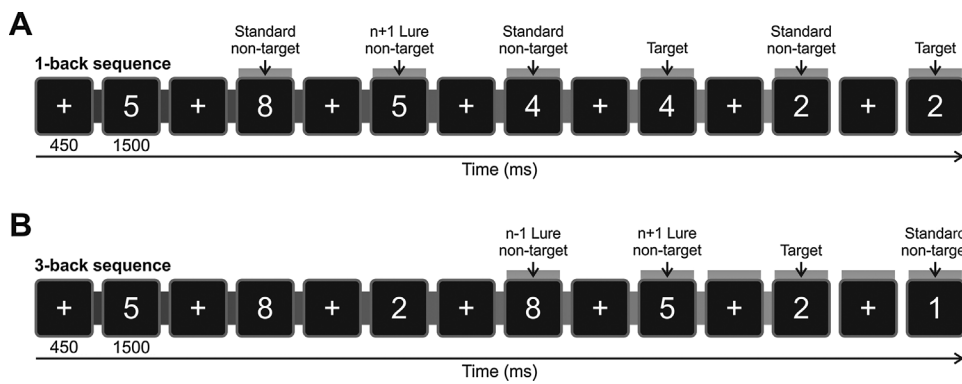


Fig. 1. Illustration of the n-back task. **A.** Schematic example of the first part of a 1-back sequence where target, standard non-target, and n + 1 lure non-target trials are shown. **B.** Schematic example of the first part of a 3-back sequence where target, standard non-target, n – 1 lure non-target, and n + 1 lure non-target trials are shown. In the present visual n-back task, numbers from 1 to 9 were presented one at a time in the middle of a computer screen. The trial began with a fixation point for 450 ms, followed by the number shown for 1500 ms. The stimulus asynchrony was fixed to 1950 ms. Participants had to press the ‘yes’ button when the trial was a target, which means that the number was the same as the previous number (1-back task) or the number presented three numbers before (3-back task). For the other, non-target numbers that included standard,

n – 1 lure, and n + 1 lure non-target trials, the participants had to press the ‘no’ button.

free correctly responded trials were included in the analyses.

2.6. EEG data analyses

Regarding the ERPs, different analyses were carried out separately for each component. For the PSW and for the late negativity, the time-window was set so that it could catch each component as a whole, and to reduce the influence of previous components. For the P2, and the P3b, the time window was selected according to visual inspection of the expected time window, and it was set as ± 50 ms around the peak amplitude. The mean amplitudes of the following time windows were thus selected for the analyses: PSW (–600–0 ms), P2 (190–290 ms), P3b (330–430 ms), late negativity (800–1200 ms).

First, the mean amplitude of the PSW for target trials was submitted to an exploratory repeated measures (rm) ANOVA including the following three within-subject factors: load (two levels: 1-back vs. 3-back), anterior-posterior (three levels: anterior frontal [electrodes F3, Fz, and F4], central [C3, Cz, and C4], posterior [P3, Pz, and P4]), and laterality (three levels: left [F3, C3, and P3], middle [Fz, Cz, and Pz], and right [F4, C4, and P4]). Second, in order to reveal the effects of load and trial type on the P2 and P3b amplitude, the mean amplitudes of both components were separately submitted to rmANOVAs. The rmANOVAs included the following within-subject factors: load (two levels: 1-back vs. 3-back), trial type (two levels: target, non-target [including both standard and lure trials]) and electrodes (three levels: Fz, Cz, Pz). To unveil the specific features of lures in the 1-back and 3-back conditions, additional separate analyses per condition were performed for P2 and P3b. For the 1-back condition, the rmANOVAs included two within-subject factors: trial type (two levels: standard non-target, n + 1 lure) and electrodes (three levels: Fz, Cz, Pz). The same design with two within-subject factors was performed for the 3-back, but here the trial type had three levels (standard non-target, n + 1 lure, n – 1 lure).

To correct for possible violations of the sphericity assumption (Jennings & Wood, 1976) the Greenhouse–Geisser epsilon correction was used, and the adjusted p-values after the correction are reported. Only significant p-values that exceeded the false discovery rate (FDR) threshold were considered (Cramer et al., 2016). The Cohen’s d and f were reported as a measure of the effect sizes for the t-tests and the ANOVAs, respectively (Cohen, 1992).

3. Results

3.1. Behavioral results

The overall analyses by a 2 × 2 (load, trial type) rmANOVA on RTs revealed the expected main effect of load, $F(1,26) = 80.86, p < .001, f = 1.765$, as decision latencies were significantly faster for the 1-back than for the 3-back condition (see Table 2). The main effect of trial type was also statistically significant, $F(1,26) = 27.60, p < .001, f = 1.031$,

Table 2

Mean Performance in Reaction Times (in Milliseconds) and Accuracy Rates with Standard Deviations in the n-back Task.

			N	M	SD
Single n-back	Targets	1-back RT	27	570.30	75.56
		3-back RT	27	680.05	111.72
		1-back accuracy	27	92.49	5.58
		3-back accuracy	27	80.27	9.52
		1-back d'	27	3.43	.73
		3-back d'	27	2.07	.63
	Non-targets	n-back effect in RT	27	109.75	68.80
		n-back effect accuracy	27	12.22	8.86
		1-back RT	27	606.78	100.92
		3-back RT	27	753.08	129.60
		1-back accuracy	27	96.07	2.83
		3-back accuracy	27	82.99	10.16
All	n-back effect in RT	27	146.30	81.22	
	n-back effect accuracy	27	13.08	9.58	
	1-back RT	27	594.62	90.63	
	3-back RT	27	734.82	123.14	
	1-back accuracy	27	94.87	3.31	
	3-back accuracy	27	82.31	8.93	
		n-back effect in RT	27	140.20	76.49
		n-back effect accuracy	27	12.57	8.72

reflecting the fact that RTs were faster for target than for non-target trials. The lack of a significant interaction between load and trial type, $F(1,26) = 3.76, p = .063, f = .380$, suggested that the difference in RTs between target and non-target trials was equally distributed across loads. The results concerning accuracy rates (Table 2) as measured by d' showed higher accuracy rates for 1-back than 3-back, $F(1,26) = 78.05, p < .001, f = 1.732$.

The separate rmANOVA on RTs for non-target types in the 1-back condition (see Table 3) showed a significant main effect, $F(1,26) = 193.50, p < .001, f = 2.334$, indicating that RTs were slower for n + 1 lures compared to standard non-target trials. For the 3-back,

Table 3

Mean Performance in Reaction Times (in Milliseconds) and Accuracy Rates with Standard Deviations in the n-back Task for non-target trials.

			N	M	SD	
Single n-back	1-back	standard non-target RT	27	560.45	95.82	
		n + 1 lure RT	27	653.12	108.57	
		Standard non-target accuracy	27	98.48	2.12	
		n + 1 lure accuracy	27	93.96	3.91	
		3-back	standard non-target RT	27	696.83	123.40
			n + 1 lure RT	27	791.86	128.63
	n – 1 lure RT		27	770.55	143.15	
	standard non-target accuracy		27	96.03	3.28	
	n + 1 lure accuracy		27	73.82	14.84	
	n – 1 lure accuracy		27	79.13	13.73	

the RT analysis revealed a significant main effect of non-target type, $F(2,52) = 71.65$, $p < .001$, $f = 1.661$. Subsequent *post-hoc* comparisons showed that RTs were slower for lures compared to standard non-target trials ($n + 1$ lures vs. standard non-target: $t(26) = 11.63$, $p < .001$, $d = .780$; $n - 1$ lures vs. standard non-target: $t(26) = 8.81$, $p < .001$, $d = -.571$), and for $n + 1$ than for $n - 1$ lure trials ($t(26) = 2.52$, $p = .028$, $d = .162$).

A similar analysis on accuracy rates (correct rejections) for non-target types in the 1-back condition (Table 3) showed a significant main effect, $F(1,26) = 82.94$, $p < .001$, $f = 1.784$, indicating that accuracy was worse for $n + 1$ lures than for standard non-target trials. For the 3-back, the analyses showed a main effect for non-target type, $F(2,52) = 65.93$, $p < .001$, $f = 1.591$. The subsequent *post-hoc* comparisons indicated that accuracy was lowest for lures compared to standard non-target trials ($n + 1$ lures vs. standard non-target: $t(26) = -9.38$, $p < .001$, $d = -2.139$; $n - 1$ lures vs. standard non-target: $t(26) = -7.54$, $p < .001$, $d = -1.753$), and for $n + 1$ compared to $n - 1$ lures ($n + 1$ lures vs. $n - 1$ lures: $t(26) = -4.18$, $p < .001$, $d = -.384$).

3.2. ERP results

As shown in Fig. 2, PSW can be observed in the 1-back condition for target trials, with an onset at about 600 ms prior to the presentation, lasting until 100–150 ms post-stimulus. Regarding scalp distribution, this PSW was most prominent over left centro-posterior regions. After stimulus onset, a P2 component peaking around 240 ms with a fronto-central distribution is seen, followed by a P3b component peaking at ~380 ms with a posterior-central distribution. On visual inspection (Fig. 2), the P2 and the P3b components appear larger for the 1-back than the 3-back condition, and for target than non-target trials. However, after removing the influence of the PSW by using a new pre-

stimulus time-window (–100 to 0 ms), the load effect reversed for both components, being larger (i.e., more positive) for the 3-back compared to the 1-back condition, and for target compared to non-target trials (Fig. 3). In order to confirm these effects, the PSW (baseline: –1000 to –800 ms; time-window: –600 to 0 ms), P2 (baseline: –100 to 0 ms; time-window: 190–290 ms) and P3b (baseline: –100 to 0 ms; time-window: 330–430 ms) mean amplitudes were submitted to the corresponding rmANOVAs.

3.2.1. PSW

Concerning load differences, a significant main effect was obtained, $F(1,26) = 72.75$, $p < .001$, $f = 1.674$, revealing the presence of the PSW for 1-back compared to 3-back targets (see Fig. 2). Importantly, despite the lack of a significant load \times anterior-posterior interaction, $F(2,52) = 2.88$, $p = .090$, $f = .333$, the significant load \times laterality interaction, $F(2,52) = 7.38$, $p = .002$, $f = .533$, and load \times anterior-posterior \times laterality interaction, $F(4,104) = 3.28$, $p = .018$, $f = .355$, together with the *post-hoc* *t*-tests indicated that the largest load effect was at left centro-posterior regions (1-back vs. 3-back at C3: $t(26) = 8.54$, $p < .001$, $d = 1.582$; 1-back vs. 3-back at P3: $t(26) = 9.26$, $p < .001$, $d = 1.462$).

3.2.2. P2

Regarding load and trial type differences, the significant main effect of load, $F(1,26) = 27.52$, $p < .001$, $f = 1.028$, stemmed from the larger amplitude of the P2 for the 3-back than for the 1-back (see Fig. 3). Similarly, a significant main effect of trial type, $F(2,52) = 20.09$, $p < .001$, $f = .879$, confirmed a larger amplitude for target than for non-target trials.

Importantly, the significant load \times trial type interaction, load \times trial type: $F(2,52) = 11.98$, $p = .002$, $f = .678$, and posterior *t*-test comparisons showed that the trial type effect remained significant only in

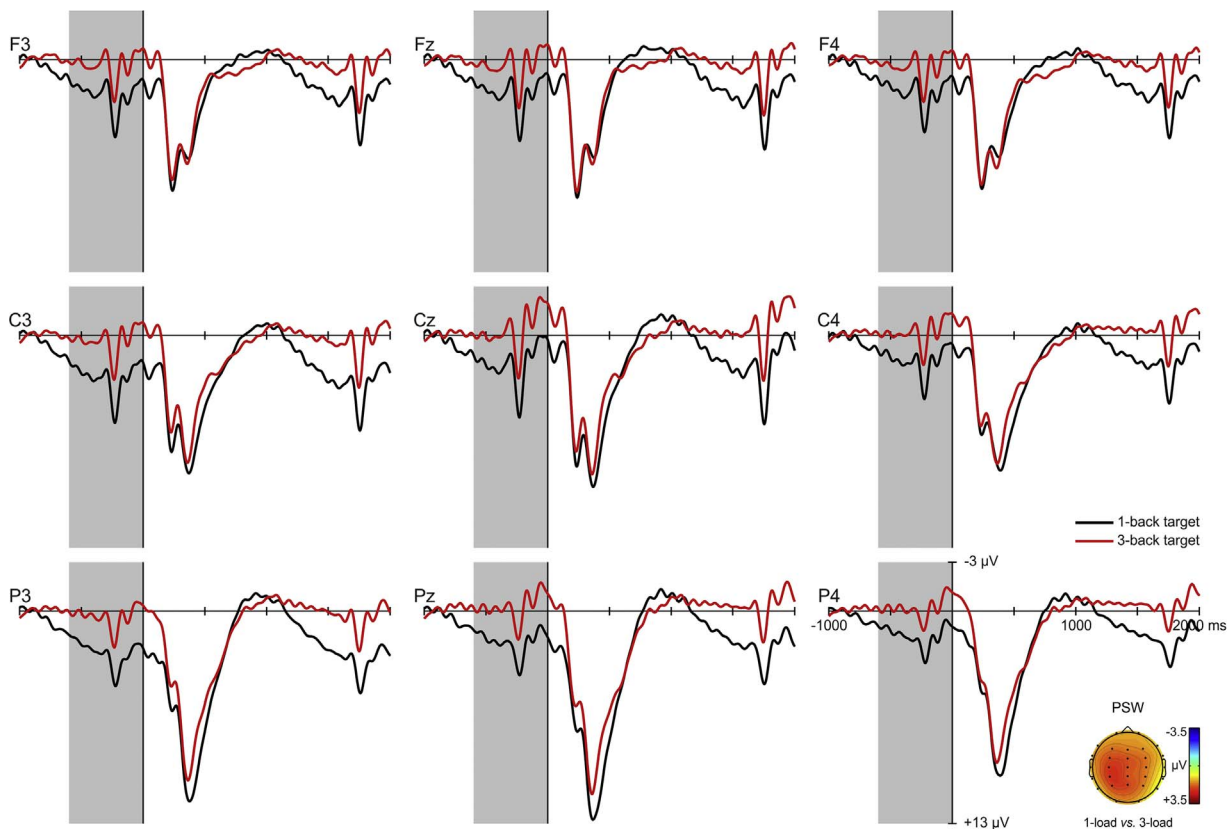


Fig. 2. ERP responses with a baseline from –1000 to –800 ms. Grand average ERPs for the 1-back and 3-back from nine electrode locations. A clear PSW prior to the stimulus was observed for 1-back. After stimulus presentation, the P2 appeared, being followed by the P3b. Scalp distributions for 1-back vs. 3-back subtractions for the PSW (time-window: –600 to 0 ms).

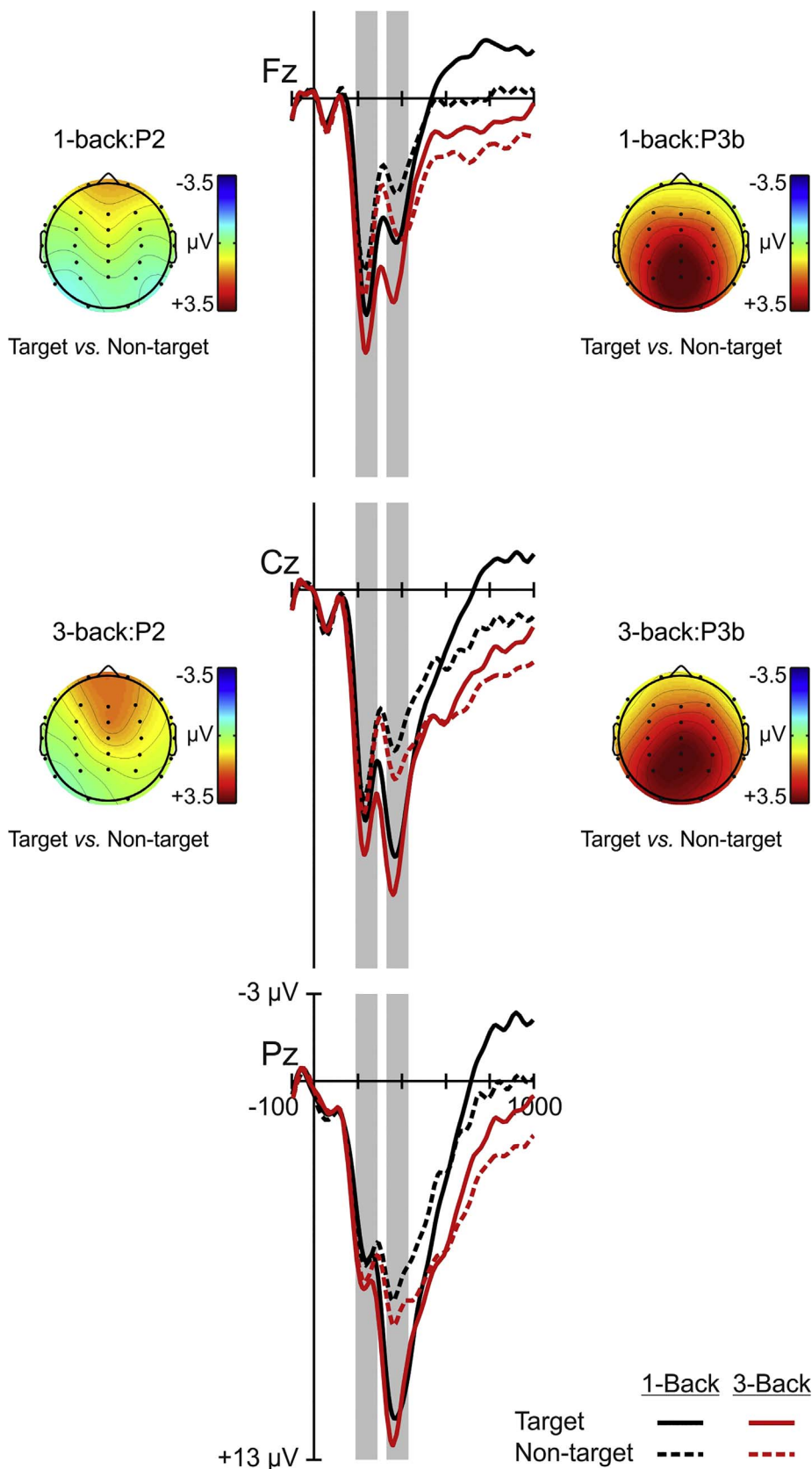


Fig. 3. ERP response with a baseline from -100 to 0 ms. Grand average ERPs for the 1-back and 3-back, and for target and non-target stimuli from central electrode locations. P2 and P3b responses were observed for both 1-back and 3-back trials. Scalp distributions for 1-back vs. 3-back subtractions for target and non-target stimuli, and for the P2 (time-window: 190 – 290 ms); and P3b (time-window: 330 – 430 ms).

the 3-back condition (mean 3-back target vs. non-target: $t(26) = 7.11$, $p < .001$, $d = .501$; mean 1-back target vs. non-target: $t(26) = 1.31$, $p = .203$, $d = .135$), being similar across the central electrodes (load \times electrodes: $F(2,52) = 2.61$, $p = .08$, $f = .316$; load \times trial

type \times electrodes: $F(2,52) = .78$, $p = .43$, $f = .173$). Furthermore, the rmANOVA together with *post-hoc* analyses revealed that the trial type effect was largest at fronto-central regions (trial type \times electrodes: $F(2,52) = 30.19$, $p < .001$, $f = 1.077$; Fz target vs. non-target: t

(26) = 5.86, $p < .001$, $d = .637$; Cz target vs. non-target: $t(26) = 4.53$, $p < .001$, $d = .350$; Pz target vs. non-target: $t(26) = -.30$; $p = .771$, $d = -.411$).

3.2.3. P3b

Following the P2, a P3b could be observed (see Fig. 3). The larger amplitude for the 3-back compared to the 1-back condition was confirmed by the main effect of load, $F(1,26) = 16.66$, $p < .001$, $f = .801$. Similarly, the significant main effect of trial type, $F(1,26) = 107.70$, $p < .001$, $f = 2.038$, reflected the larger amplitude for the target than for the non-target trials. The non-significant load \times trial type interaction, $F(1,26) = .014$, $p = .907$, $f = .032$, suggested that the larger amplitude for 3-back than for 1-back remained stable across trial types. The highest difference between 3-back and 1-back was encountered at fronto-central sites, as was indicated by the load \times electrode interaction, $F(2,52) = 9.69$, $p = .001$, $f = .611$, and the *post-hoc* comparisons (Fz 3-back vs. 1-back: $t(26) = 5.62$, $p < .001$, $d = .543$; Cz 3-back vs. 1-back: $t(26) = 3.12$, $p = .004$, $d = .324$; Pz 3-back vs. 1-back: $t(26) = 2.43$, $p = .022$, $d = .199$). On the other hand, the largest difference between target and non-target trials was observed at the centro-posterior regions (trial type \times electrode: $F(2,52) = 36.50$, $p < .001$, $f = 1.185$; load \times trial type \times electrode: $F(2,52) = 3.06$, $p = .069$, $f = .342$; Fz target vs. non-target: $t(26) = 5.69$, $p < .001$, $d = .592$; Cz target vs. non-target: $t(26) = 10.02$, $p < .001$, $d = 1.103$; Pz target vs. non-target: $t(26) = 11.81$, $p < .001$, $d = 1.092$).

3.2.4. Decomposing 1-back non-target trials

In order to delineate the lure effects on the 1-back and 3-back conditions, non-target trials were decomposed (see Fig. 4). Accordingly, specific rmANOVAs were conducted on the P2, P3b and late negativity, separately for the 1-back and 3-back conditions.

For the P2 in the 1-back condition, a significant main effect of non-target trial type was obtained in the rmANOVA, $F(1,26) = 7.80$, $p = .009$, $f = .554$, reflecting a reduction in the P2 amplitude for $n + 1$ lures compared to standard non-target trials. The non-significant trial type \times electrode interaction, $F(2,52) = 1.13$, $p = .330$, $f = .209$, indicated that the effect was consistent across electrodes.

Concerning the P3b, we did not find a significant main effect of trial type, $F(1,26) = .019$, $p = .890$, $f = .032$, or trial type \times electrode interaction, $F(2,52) = .45$, $p = .560$, $f = .131$, indicating similar P3b amplitude as well as distribution between $n + 1$ lure and standard non-target trials.

Interestingly, a clear delayed late negativity appeared for $n + 1$ lure trials when compared to non-target trials between 800–1200 ms. The mean amplitude of the negativity in this time-window was submitted to a rmANOVA, and the main effect of trial type was significant, $F(1,26) = 5.59$, $p = .022$, $f = .473$. This negativity was similar across electrode positions, as the trial type \times electrode interaction was not significant, $F(2,52) = .84$, $p = .395$, $f = .179$.

3.2.5. Decomposing 3-back non-target trials

With respect to the 3-back, we conducted the same analyses as for 1-back (see above). For the P2 amplitude, the results showed a significant main effect of trial type, $F(2,52) = 3.72$, $p = .034$, $f = .378$. The interaction between trial type and electrode, $F(4,104) = 2.74$, $p = .055$, $f = .325$, was not significant, as the trial type exerted similar effects across electrodes. *Post-hoc* analyses detected larger P2 amplitude for standard non-target trials compared to $n + 1$ lures, $t(26) = 3.10$, $p = .005$, $d = .318$. However, no significant differences were observed between non-target trials and $n - 1$ lures, $t(26) = .91$, $p = .369$, $d = .117$, or between $n + 1$ and $n - 1$ lures, $t(26) = -1.65$, $p = .111$, $d = -.207$.

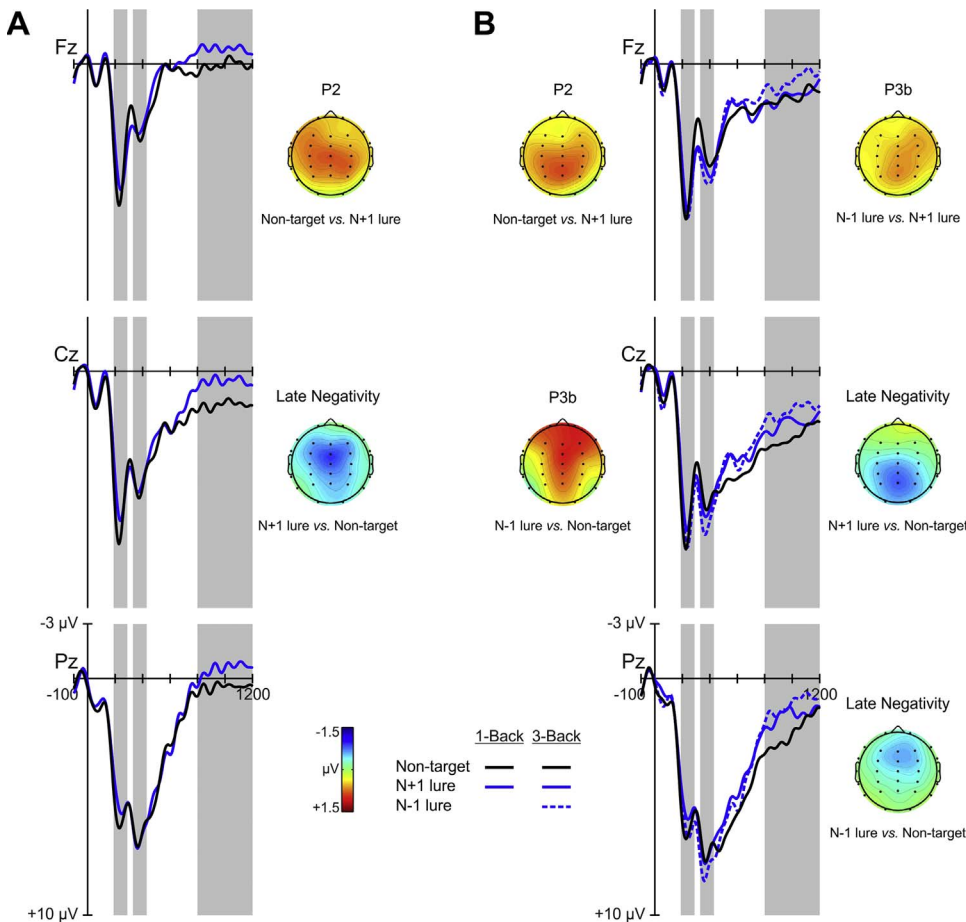


Fig. 4. ERP response with a baseline from -100 to 0 ms. A. Grand average ERPs at central electrodes, for non-target trial types (standard non-target, $n - 1$ lure) in the 1-back condition. B. Grand average ERPs at central electrodes, for non-target trial types (standard non-target, $n + 1$ lure, $n - 1$ lure) in the 3-back condition. P2 (time-window: 190–290 ms) and P3b (time-window: 330–430 ms) responses were observed for all types of non-target trials, in both 1-back and 3-back. On the other hand, late negativity (time-window: 800–1200 ms) could be detected for lures in both 1-back and 3-back. Scalp distributions were plotted on these time-windows only for the significant effects. Concerning the 1-back, subtractions for the P2 (standard non-target vs. $n + 1$ lure) and late negativity ($n + 1$ lure vs. standard non-target) were conducted. Regarding the 3-back, subtractions for the P2 (standard non-target vs. $n + 1$ lure), P3b ($n - 1$ lure vs. standard non-target; $n - 1$ lure vs. $n + 1$ lure), and late negativity ($n + 1$ lure vs. standard non-target; $n - 1$ lure vs. standard non-target) were performed.

Regarding P3b, the rmANOVA showed a significant main effect of trial type, $F(2,52) = 7.47$, $p = .002$, $f = .536$, confirming amplitude differences between non-target trials. The non-significant trial type \times electrode interaction, $F(4,104) = 1.34$, $p = .268$, $f = .227$, suggested that the observed differences were similar across electrode locations. *Post-hoc* pairwise comparisons revealed that the P3b amplitude was larger for $n - 1$ lures compared to other non-target trials ($n - 1$ lure vs. standard non-target: $t(26) = 3.55$, $p = .001$, $d = .382$; $n - 1$ lure vs. $n + 1$ lure: $t(26) = 2.11$, $p = .018$, $d = .235$). The difference between $n + 1$ lures and standard non-target trials was non-significant, $t(26) = 1.51$, $p = .114$, $d = .150$.

Concerning the amplitude of the late negativity (800–1200 ms), a significant main effect of trial type was encountered, $F(2,52) = 6.85$, $p = .004$, $f = .514$. The significant interaction between trial type and electrode, $F(4,104) = 5.46$, $p = .003$, $f = .174$, together with *post-hoc* comparisons, indicated that the negativity was higher for lures compared to standard non-target trials ($n - 1$ lures vs. standard non-target: $t(26) = 3.64$, $p = .001$, $d = .513$; $n + 1$ lures vs. standard non-targets: $t(26) = -2.52$, $p = -.018$, $d = -.267$). However, no difference was found between the $n + 1$ vs. $n - 1$ lures, $t(26) = 1.48$, $p = .151$, $d = .231$.

3.3. Correlations between behavioral and ERP data

To obtain further evidence of the role of PSW in active WM maintenance, we examined whether its amplitude would predict task performance specifically in the 1-back task. Indeed, PSW amplitude at the P3 electrode in the 1-back task correlated with RTs to correctly responded targets ($r = -.403$, $p < .05$; see Fig. 5A). No significant correlation between PSW and RT was observed for the 3-back condition ($r = -.051$) or for the 3-back vs. 1-back load effect ($r = -.273$). The difference between the PSW – RT correlation coefficients in the 1-back and 3-back tasks, measured by Fisher's r -to- z transform, was nevertheless only marginally significant ($p = .09$). PSW amplitude at P3 predicted the subsequent P2 amplitude at Fz electrode ($r = .396$, $p < .05$; see Fig. 5B). The correlation with RT was significant for P2 amplitude at Fz electrode ($r = -.429$, $p < .05$) but not for P3b at Pz ($r = -.375$, $p = .054$), but neither one of these correlations was significant when PSW amplitude was partialled out in the correlation ($p > .09$). In a similar vein, the difference between the two correlation coefficients was not significant ($p = .92$). Finally, no significant correlation was observed between d' and ERP response amplitudes ($p > .09$). The high accuracy (easiness) in the 1-back condition diminished the chances for finding such a correlation in the response accuracy. Thus the link between ERPs and task performance was limited to RTs and driven by PSW only.

4. Discussion

The present ERP experiment aimed to shed light on the neural correlates of active maintenance and management of interference during WM updating. Bailey et al. (2016) recently demonstrated that the 1-back condition triggers a slow-wave pre-stimulus ERP response, arguing that this reflects the active maintenance of the previous stimulus in WM. In line with their findings, we found a pre-stimulus PSW in the 1-back condition. Furthermore, we demonstrated that PSW is associated with performance in the 1-back task, providing the first direct evidence for its role in active maintenance of the task-relevant stimulus in WM (Fig. 2). The PSW was followed by P2 and P3b components that are typically reported in ERP studies examining updating (Fig. 2). As previous studies have demonstrated (Gevins et al., 1996; Anurova et al., 2005), these responses were clearly enhanced for target stimuli (Fig. 3). In the initial analysis with a long pre-stimulus baseline (Fig. 2), visual inspection did not show a load effect in P2 and P3 responses. Actually, these effects appeared to be larger in the 1-back task. With a shorter pre-stimulus baseline that accounts for PSW, load-

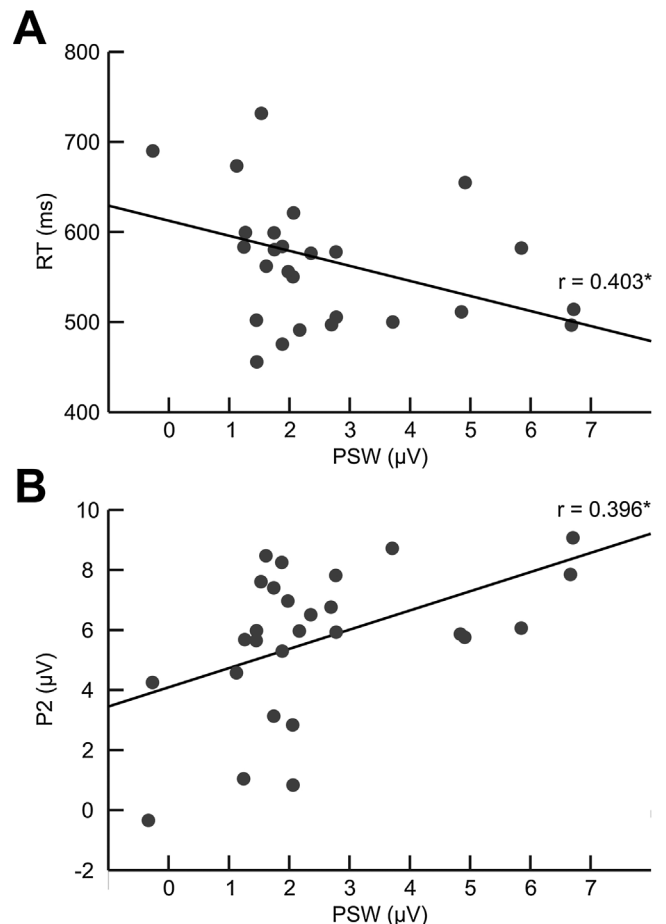


Fig. 5. Scatter plots of the significant correlations observed for the amplitude of the PSW during the 1-back task. A. Correlation between the PSW amplitude at P3 electrode and RT for correctly responded targets. B. Correlations between the amplitudes of the PSW at P3 electrode and P2/P3 at Fz electrode.

dependent response in the P2/P3 was, however, observed (Fig. 3). Our study is the first to suggest that PSW should be considered when examining the load-dependent responses. This issue is highly relevant for analyzing and interpreting similar studies. Besides, according to our findings with lure stimuli, these responses are also modulated by interference when the WM load is high. This was observed as modulation of P3 response by the $n + 1$ lures in 3-back, perhaps reflecting increased allocation of attention in a situation where the lure elicits uncertainty for decision-making. Interestingly, in the 1-back task the $n + 1$ lures modulated mostly sustained activity at long latencies, reflecting post-decision processes that could be related to conflict detection or increasing response monitoring (Folstein & Van Petten, 2008; Ford, Roth, Mobs, Hopkins, & Kopell, 1979; Pelosi, Hayward, & Blumhardt, 1995). These late responses have been observed in prior WM studies (Marchand et al., 2006; Evans & Federmeier, 2007; Lefebvre et al., 2005; Rämä et al., 2000; Bailey et al., 2016), but to our knowledge this is the first study demonstrating interference effects at these latencies. It is an open issue why the lures in 1-back did not affect P3 responses albeit they exerted a clear behavioral effect. This may have to do with some fundamental differences in decision-making in 1-back vs 3-back. The former one operates with a much narrower scope affected by maintenance and stimulus recency effects, while the latter one calls for memory retrieval.

Prior studies have reported that persistent sensory cortical activity during WM discrimination tasks provides an index of efficient selectivity of contents stored in WM (see Fukuda et al., 2010; Luria et al., 2016). PSW has been linked to maintenance of WM in delayed

matching-to-sample tasks, which dissociate between encoding, maintenance, and retrieval processes (e.g., Ruchkin et al., 1990; Rösler & Heil, 1991; Ruchkin et al., 1995; Mecklinger & Müller, 1996; Mecklinger & Pfeifer, 1996), while only the study by Bailey et al. (2016) has previously linked it to maintenance during a WM updating task. In line with Bailey et al. (2016), anticipatory PSW response was observed both in the 3-back and 1-back condition, but it was clearly enhanced in the 1-back condition during which maintenance of stimulus information associated with comparison process is much easier than during the 3-back condition where higher updating demands make the retrieval of the target item more difficult. To verify and extend these findings, we used a n-back setup that avoided familiarity-based decisions (e.g., Kane et al., 2007) through inclusion of lure trials, and correlated the slow wave amplitudes with decision latencies. We found that PSW amplitude predicted response times to correct responses only in the 1-back task (accuracy rates could not be related to PSW, as the participants made very few errors on 1-back). When PSW amplitude was accounted for, the subsequent ERP components were no longer associated with the task performance. Altogether, these findings provide further support to the view on PSW as a neural signature for active maintenance of information in WM.

It is worth noting that the scalp distribution of PSW was somewhat different between our study and that of Bailey et al. (2016). While we observed a relatively widespread distribution covering anterior and posterior scalp areas, Bailey and colleagues reported clearly more frontally-weighted scalp distributions. As they used common average reference, we tested the influence of the reference electrode on PSW scalp distribution. Re-referring to common average did not, however, lead to changes in the frontal scalp distribution in our study. It therefore seems that these differences in scalp distribution may relate to the differences in the experimental tasks. In our experiment, the n-back task was performed with digits, while Bailey et al. (2016) used letters as stimuli. These two tasks relying on different content representations (see Eriksson, Vogel, Lansner, Bergström, & Nyberg, 2015) could elicit activations in partly different brain areas, numbers being processed mostly in the posterior parietal cortex (e.g., Hubbard, Piazza, Pinel, & Dehaene, 2005; Bueti & Walsh, 2009) and purely verbal stimuli showing particularly prominent activations in the prefrontal and temporal areas (Veltman, Rombouts, & Dolan, 2003; Owen et al., 2005). It should be noted, however, that this interpretation remains speculative as no direct comparison between n-back n-back tasks with different types of stimulation was made and ERP studies do not have the optimal spatial resolution to detect the exact sources of PSW response. This idea is in agreement also with the results from another slow pre-stimulus component, the Stimulus Preceding Negativity (SPN), which also shows differences in topographies and cortical sources across studies that are most probably related to modality (nature of the anticipated stimuli) and task-specific effects (Kotani et al., 2015; Morís, Luque, & Rodríguez-Fornells, 2013; Hackley, Valle-Inclán, Masaki, & Hebert, 2014; Van der Molen et al., 2013).

PSW may further cause delay in the return to the baseline by decreasing the slope of its negative-going part (García-Larrea & Cezanne-Bert, 1998). Our findings suggest that when this delay is accounted for in the pre-stimulus period, the load increase from low load to high load is observed in P2 and P3b amplitudes (see Fig. 2 vs. 3). Thus, slow anticipatory response to maintenance may possibly explain some of the discrepancies in load-dependent changes in P2 and P3b in earlier studies. In fact, by going through the previous studies afterwards, we noted that in earlier studies several of those reporting larger effects with higher load have used -100 ms baseline (Lefebvre et al., 2005; Evans & Federmeier, 2007; Chen et al., 2008; Chen & Mitra, 2009; Marchand et al., 2006), while others reporting larger effects in 1-back tasks have used -200 ms baseline (Rämä et al., 2000; Dong et al., 2015; Bailey et al., 2016).

As noted in Introduction, the n-back paradigm is one of the most widely used experimental tasks in the research of neurocognitive

underpinnings of WM. The number of studies examining the adaptations of brain responses associated with training has accumulated rapidly (see Constantinidis & Klingberg, 2016). However, there has been a shortage of measures allowing to map training-related changes in brain responses to specific WM component processes. Verification of a neural marker for active maintenance in the n-back task can increase the usefulness of this task paradigm, for instance, in revealing the alterations of WM function associated with clinical conditions or training (see Müller & Knight, 2006; Au et al., 2015; Soveri et al., 2017). It should be noted, however, that this was only the second study showing the link between PSW and WM maintenance. More research is therefore needed to understand the neurocognitive basis of PSW. For instance, MEG would likely have high enough spatial resolution to clarify whether PSW is related to higher-level WM component processes associated with maintenance, or to persistent modulation of representational contents by influencing WM “slave systems” (see Eriksson et al., 2015). Finally, our study leaves it open whether PSW is also associated with response accuracy as 1-back tasks were performed with very high accuracy. However, if PSW amplitude reflects the success in maintaining the previous stimulus in WM, correlation between brain response and task performance may not necessarily be observed for accuracy as one can respond correctly by performing a memory search that is triggered by the next stimulus.

Our findings also revealed novel aspects concerning interference effects in WM. As in previous studies, lures delayed stimulus matching and increased the number of errors. In this respect, speed and difficulty of the lure processing was approximately at the level of target processing. Behaviorally, the lures showed largely similar influences on task performance in the 1-back and in 3-back condition, but our ERP findings suggest that two distinct neurophysiological processes underlie these behaviorally seemingly similar effects. High memory load resulted to interference at the earlier response selection and comparison stage, while at low memory load the significant interference effects shifted to post-decision processes.

Both our behavioral and ERP findings suggested that $n + 1$ lures caused higher interference than $n - 1$ lures in this group of participants and in this particular task. So far, the evidence for differential lure effects has been scarce. Szmalec et al. (2011) reported more slowing in RTs for $n - 1$ lures than $n + 1$ lures (Szmalec et al., 2011), but Ralph (2014) has more recently argued that the strength of $n - 1$ vs. $n + 1$ lure effects depends on strategies that the participants use, and those may vary according to the specific type of n-back task. Ralph (2014) argued that proactive strategies that weight the target item would cause higher interference and slower response times for $n + 1$ lures, and this corresponds to our empirical findings. Ralph (2014) also hypothesized that the lure effect would be reflected in brain responses at 200 ms latencies (N2 response in particular), as is the case with conflict monitoring in Flanker and Stroop tasks (Yeung, Bogacz, Holroyd, & Cohen, 2004; Yeung, Bogacz, Holroyd, Nieuwenhuis, & Cohen, 2007). We did observe lure effects at these latencies, albeit on P2 and P3 responses. Importantly, the central distribution of the late negativity observed in lure trials might fit with the scalp distribution of N2-N3 conflict-related negative components associated to the detection of a mismatch between the stimulus features presented (Yeung et al., 2004), or between the stimulus and some previously formed internal template (Folstein & Van Petten, 2008). These frontal-central negativities appear when participants are engaged in some sort of a comparison processes (response monitoring). Interestingly, N2 frontal effects have been also observed when visually presented stimulus arrays do not match participant's expectancies created by a previous cue (for example, they have been observed using the Sternberg paradigm, when a participant decides if a probe item matches or not with a small number of items presented before). A larger frontal N2 effect peaking around 270–340 ms has been elicited by probe items that did not match the previous item set (Ford et al., 1979; Pelosi et al., 1995). Although this issue remains to be studied further, the late negative component that we observed could be

related to post-response monitoring strategies engaged in increased conflict or interference.

5. Conclusions

Our findings demonstrate that active maintenance of information during a WM updating task elicits a cascade of brain responses starting from pre-stimulus PSW and being followed by P2 and P3b responses, with the latter two responses being influenced by PSW amplitude. This study also provided the first direct evidence of the contribution of this response to efficient WM behavioral performance: higher PSW amplitude was associated with improved task performance as indexed by shorter response times to correctly responded targets in the 1-back task. Altogether, this evidence supports the view that the stimulus-preceding PSW signals active maintenance of the previous task-relevant stimulus. This neuronal marker of WM maintenance has several potential implications, for example in investigating how neuropsychiatric conditions or systematic training alter WM function. If PSW is not accounted for, P2 and P3b response amplitudes seem to be smaller for higher WM load. This effect, however, appears to be reversed and thus showing a traditional load effect when the PSW amplitude at the baseline is taken into account. Finally, we provide the first evidence that ERPs following lure stimuli show differential effects at high vs. low load levels, suggesting two distinct neuronal correlates for interference during WM updating.

Acknowledgements

The study was supported by funds from the Åbo Akademi University Endowment (BrainTrain project) and the Academy of Finland (grant #260276) awarded to ML. Special thanks to all members of BrainTrain research group for their valuable comments.

References

- Albouy, P., Weiss, A., Baillet, S., & Zatorre, R. J. (2017). Selective entrainment of theta oscillations in the dorsal stream causally enhances auditory working memory performance. *Neuron*, *94*, 193–206.
- Anurova, I., Artchakov, D., Korvenoja, A., Ilmoniemi, R. J., Aronen, H. J., & Carlson, S. (2005). Cortical generators of slow evoked responses elicited by spatial and non-spatial auditory working memory tasks. *Clinical Neurophysiology*, *116*, 1644–1654.
- Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuhl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin and Review*, *22*, 366–377.
- Bailey, K., Mlynarczyk, G., & West, R. (2016). Slow wave activity related to working memory maintenance in the N-Back task. *Journal of Psychophysiology*, *30*, 141–154.
- Botto, M., Basso, D., Ferrari, M., & Palladino, P. (2014). When working memory updating requires updating: Analysis of serial position in a running memory task. *Acta Psychologica (Amsterdam)*, *148*, 123–129.
- Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time space, number and other magnitudes. *Philosophical Transactions of the Royal Society of London: Series B*, *364*, 1831–1840.
- Bunting, M., Cowan, N., & Saults, J. S. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology*, *59*, 1691–1700.
- Burgess, G. C., Gray, J. R., Conway, A. R., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General*, *140*, 674–692.
- Chapman, R. M., & Bragdon, H. R. (1964). Evoked Responses to Visual and Non-Numerical Visual Stimuli while Problem Solving. *Nature*, *203*, 1155–1157.
- Chen, Y. N., & Mitra, S. (2009). The spatial-verbal difference in the n-back task: An ERP study. *Acta Neurologica Taiwanica*, *18*, 170–179.
- Chen, Y. N., Mitra, S., & Schlaghecken, F. (2008). Sub-processes of working memory in the N-back task: An investigation using ERPs. *Clinical Neurophysiology*, *119*, 1546–1559.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.
- Constantinidis, C., & Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nature Reviews Neuroscience*, *17*, 438–449.
- Cramer, A. O., van Ravenzwaaij, D., Matzke, D., Steingrover, H., Wetzels, R., Grasman, R. P., & Wagenmakers, E. J. (2016). Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychonomic Bulletin & Review*, *23*, 640–647.
- Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology*, *115*, 732–744.
- Daffner, K. R., Chong, H., Sun, X., Tarbi, E. C., Riis, J. L., McGinnis, S. M., & Holcomb, P. J. (2011). Mechanisms underlying age- and performance-related differences in working memory. *Journal of Cognitive Neuroscience*, *23*, 1298–1314.
- Dien, J., Spencer, K. M., & Donchin, E. (2004). Parsing the late positive complex: Mental chronometry and the ERP components that inhabit the neighborhood of the P300. *Psychophysiology*, *41*, 665–678.
- Donchin, E., Miller, G. A., & Farwell, L. A. (1986). The endogenous components of the event-related potential—A diagnostic tool? *Progress in Brain Research*, *70*, 87–102.
- Dong, S., Reeder, L. M., Yao, Y., Liu, Y., & Chen, F. (2015). Individual differences in working memory capacity are reflected in different ERP and EEG patterns to task difficulty. *Brain Research*, *1616*, 146–156.
- Ecker, U. K., Lewandowsky, S., Oberauer, K., & Chee, A. E. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 170–189.
- Ecker, U. K., Maybery, M., & Zimmer, H. D. (2013). Binding of intrinsic and extrinsic features in working memory. *Journal of Experimental Psychology: General*, *142*, 218–234.
- Eriksson, J., Vogel, E. K., Lansner, A., Bergström, F., & Nyberg, L. (2015). Neurocognitive architecture of working memory. *Neuron*, *88*, 33–46.
- Evans, K. M., & Federmeier, K. D. (2007). The memory that's right and the memory that's left: Event-related potentials reveal hemispheric asymmetries in the encoding and retention of verbal information. *Neuropsychologia*, *45*, 1777–1790.
- Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP 'old/new' effects: Memory strength and decisional factor(s). *Neuropsychologia*, *40*, 2288–2304.
- 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.
- Ford, J. M., Roth, W. T., Mobs, R. C., Hopkins, W. F., & Kopell, B. S. (1979). Event-related potentials recorded from young and old adults during a memory retrieval task. *Electroencephalography and Clinical Neurophysiology*, *47*, 450–459.
- Fukuda, K., Awh, E., & Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology*, *20*, 177–182.
- Fuster, J. M. (2000). Prefrontal neurons in networks of executive memory. *Brain Research Bulletin*, *52*, 331–336.
- García-Larrea, L., & Cezanne-Bert, G. (1998). P3, positive slow wave and working memory load: A study on the functional correlates of slow wave activity. *Electroencephalography and Clinical Neurophysiology*, *108*, 260–273.
- Gevins, A., Smith, M. E., Le, J., Leong, H., Bennett, J., Martin, N., ... Whitfield, S. (1996). High resolution evoked potential imaging of the cortical dynamics of human working memory. *Electroencephalography and Clinical Neurophysiology*, *98*, 327–348.
- Hackley, S. A., Valle-Inclán, F., Masaki, H., & Hebert, K. (2014). Stimulus-preceding negativity (SPN) and attention to rewards. In G. R. Magnuson (Ed.), *Cognitive Electrophysiology of Attention* (pp. 216–225). San Diego, CA: Academic Press.
- Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, *6*, 435–448.
- Jaeggi, S. M., Buschkuhl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*, 394–412.
- Jennings, J. R., & Wood, C. C. (1976). Letter: The epsilon-adjustment procedure for repeated-measures analyses of variance. *Psychophysiology*, *13*, 277–278.
- Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the N-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *33*, 615–622.
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*, 352–358.
- Klaver, P., Talsma, D., Wijers, A. A., Heinze, H. J., & Mulder, G. (1999). An event-related brain potential correlate of visual short-term memory. *NeuroReport*, *13*, 2001–2005.
- Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*, 557–577.
- Kotani, Y., Ohgami, Y., Ishiwata, T., Arai, J. I., Kiryu, S., & Inoue, Y. (2015). Source analysis of stimulus-preceding negativity constrained by functional magnetic resonance imaging. *Biological Psychology*, *111*, 53–64.
- Kramer, A. F., Wickens, C. D., & Donchin, E. (1987). Processing of stimulus properties: Evidence for dual-task integrality. *Journal of Experimental Psychology: Human Perception & Performance*, *11*, 393–408.
- Lefebvre, C. D., Marchand, Y., Eskes, G. A., & Connolly, J. F. (2005). Assessment of working memory abilities using an event-related brain potential (ERP) – Compatible digit span backward task. *Clinical Neurophysiology*, *116*, 1665–1680.
- Lenartowicz, A., Escobedo-Quiroz, R., & Cohen, J. D. (2010). Updating of context in working memory: An event-related potential study. *Cognitive, Affective and Behavioral Neuroscience*, *10*, 298–315.
- Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neuroscience and Biobehavioral Reviews*, *62*, 100–108.
- Luu, P., Caggiano, D. M., Geyer, A., Lewis, J., Cohn, J., & Tucker, D. M. (2014). Time-course of cortical networks involved in working memory. *Frontiers in Human Neuroscience*, *8*, 4.
- Marchand, Y., Lefebvre, C. D., & Connolly, J. F. (2006). Correlating digit span performance and event-related potentials to assess working memory. *International Journal of Psychophysiology*, *62*, 280–289.
- McEvoy, L. K., Smith, M. E., & Gevins, A. (1998). Dynamic cortical networks of verbal and spatial working memory: Effects of memory load and task practice. *Cerebral Cortex*, *8*, 563–574.
- Mecklinger, A., & Müller, N. (1996). Dissociations in the processing of what and where information in working memory: An event-related potential analysis. *Journal of Cognitive Neuroscience*, *8*, 453–473.
- Mecklinger, A., & Pfeifer, E. (1996). Event-related potentials reveal topographical and temporal distinct neuronal activation patterns for spatial and object working memory. *Brain Research. Cognitive Brain Research*, *4*, 211–224.

- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex Frontal Lobe tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.
- Morís, J., Luque, D., & Rodríguez-Fornells, A. (2013). Learning-induced modulations of the stimulus-preceding negativity. *Psychophysiology*, *50*, 931–939.
- Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, *81*, 111–121.
- Muller, N. G., & Knight, R. T. (2006). The functional neuroanatomy of working memory: Contributions of human brain lesion studies. *Neuroscience*, *139*, 51–58.
- Oelhafen, S., Nikolaidis, A., Padovani, T., Blaser, D., Koenig, T., & Perrig, W. J. (2013). Increased parietal activity after training of interference control. *Neuropsychologia*, *51*, 2781–2790.
- Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*, 46–59.
- Pelosi, L., Hayward, M., & Blumhardt, L. D. (1995). Is memory-scanning time in the Sternberg paradigm reflected in the latency of event-related potentials? *Electroencephalography and Clinical Neurophysiology*, *96*, 44–55.
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128–2148.
- Postle, B. R., Druzgal, T. J., & D'Esposito, M. (2003). Seeking the neural substrates of visual working memory storage. *Cortex*, *39*, 927–946.
- Rämä, P., Paavilainen, L., Anourova, I., Alho, K., Reinikainen, K., Sipilä, S., & Carlson, S. (2000). Modulation of slow brain potentials by working memory load in spatial and nonspatial auditory tasks. *Neuropsychologia*, *38*, 913–922.
- Rösler, F., & Heil, M. (1991). Toward a functional categorization of slow waves: Taking into account past and future events. *Psychophysiology*, *28*, 344–358.
- Ralph, J. (2014). *Statistical manipulation and control strategies of the n-back task*. PhD thesis. Troy, New York: Rensselaer Polytechnic Institute.
- Ruchkin, D. S., Johnson, R., Jr., Canoune, H., & Ritter, W. (1990). Short-term memory storage and retention: An event-related brain potential study. *Electroencephalography and Clinical Neurophysiology*, *76*, 419–439.
- Ruchkin, D. S., Johnson, R., Jr., Grafman, J., Canoune, H., & Ritter, W. (1992). Distinctions and similarities among working memory processes: An event-related potential study. *Brain Research. Cognitive Brain Research*, *1*, 53–66.
- Ruchkin, D. S., Canoune, H. L., Johnson, R., Jr., & Ritter, W. (1995). Working memory and preparation elicit different patterns of slow wave event-related brain potentials. *Psychophysiology*, *32*, 399–410.
- Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*, 251–257.
- Schmiedek, F., Li, S. C., & Lindenberger, U. (2009). Interference and facilitation in spatial working memory: Age-associated differences in lure effects in the n-back paradigm. *Psychology and Aging*, *24*, 203–210.
- Schmiedek, F., Lövdén, M., & Lindenberger, U. (2014). A task is a task is a task: putting complex span, n-back, and other working memory indicators in psychometric context. *Frontiers in Psychology*, *5*, 1475.
- Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin and Review*, *24*, 1077–1096.
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*, 137–149.
- Stroux, D., Shushakova, A., Geburek-Höfer, A. J., Ohrmann, P., Rist, F., & Pedersen, A. (2016). Deficient interference control during working memory updating in adults with ADHD: An event-related potential study. *Clinical Neurophysiology*, *127*, 452–463.
- Szmalc, A., Verbruggen, F., Vandierendonck, A., & Kemps, E. (2011). Control of interference during working memory updating. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 37–151.
- Van der Molen, M. J., Poppelaars, E. S., Van Hartingsveldt, C. T., Harrewijn, A., Gunther Moor, B., & Westenberg, P. M. (2013). Fear of negative evaluation modulates electrocortical and behavioral responses when anticipating social evaluative feedback. *Frontiers in Human Neuroscience*, *7*, 936.
- Veltman, D. J., Rombouts, S. A., & Dolan, R. J. (2003). Maintenance versus manipulation in verbal working memory revisited: an fMRI study. *Neuroimage*, *18*, 247–256.
- Yeung, N., Bogacz, R., Holroyd, C. B., & Cohen, J. D. (2004). Detection of synchronized oscillations in the electroencephalogram: an evaluation of methods. *Psychophysiology*, *41*, 822–832.
- Yeung, N., Bogacz, R., Holroyd, C. B., Nieuwenhuis, S., & Cohen, J. D. (2007). Theta phase resetting and the error-related negativity. *Psychophysiology*, *44*, 39–49.