Biological Psychology 123 (2017) 47–61

Contents lists available at ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

Relationship of P3b single-trial latencies and response times in one, two, and three-stimulus oddball tasks

Matthew M. Walsh[a,*], Glenn Gunzelmann[b], John R. Anderson[c]

[a] TiER1 Performance Solutions, Covington, KY 41011, United States
[b] Air Force Research Laboratory, Wright-Patterson Air Force Base, OH 45433, United States
[c] Carnegie Mellon University, Department of Psychology, Pittsburgh, PA 15213, United States

ARTICLE INFO

Article history:
Received 1 September 2016
Received in revised form
22 November 2016
Accepted 23 November 2016
Available online 25 November 2016

Keywords:
P300
EEG
Hidden semi-Markov models
Multi-voxel pattern analysis
N1

ABSTRACT

The P300 is one of the most widely studied components of the human event-related potential. According to a longstanding view, the P300, and particularly its posterior subcomponent (i.e., the P3b), is driven by stimulus categorization. Whether the P3b relates to tactical processes involved in immediate responding or strategic processes that affect future behavior remains controversial, however. It is difficult to determine whether variability in P3b latencies relates to variability in response times because of limitations in the methods currently available to quantify the latency of the P3b during single trials. In this paper, we report results from the Psychomotor Vigilance Task (PVT), the Hitchcock Radar Task, and a 3-Stimulus Oddball Task. These represent variants of the one-, two-, and three-stimulus oddball paradigms commonly used to study the P3b. The PVT requires simple detection, whereas the Hitchcock Radar Task and the 3-Stimulus Task require detection and categorization. We apply a novel technique that combines hidden semi-Markov models and multi-voxel pattern analysis (HSMM-MVPA) to data from the three experiments. HSMM-MVPA revealed a processing stage in each task corresponding to the P3b. Trial-by-trial variability in the latency of the processing stage correlated with response times in the Hitchcock Radar Task and the 3-Stimulus Task, but not the PVT. These results indicate that the P3b reflects a stimulus categorization process, and that its latency is strongly associated with response times when the stimulus must be categorized before responding. In addition to those theoretical insights, the ability to detect the onset of the P3b and other components on a single-trial basis using HSMM-MVPA opens the door for new uses of mental chronometry in cognitive neuroscience.

1. Introduction

The P3 is among the most widely studied components in event-related potential (ERP) research. It comprises a family of brain potentials including the novelty P3 and the target P3 (Sutton, Braren, Zubin, & John, 1965; for reviews, see Donchin & Coles, 1988 ; Picton, 1992; Polich, 2007). The novelty P3, or P3a, is a frontocentral positivity evoked by the onset of novel, non-target stimuli. The target P3, or P3b, has a more posterior distribution and is evoked by the onset of task-relevant stimuli that require some type of mental or motor response. In this paper, we focus on the P3b and its relationship with the timing and types of mental events culminating in behavioral responses.

A long held view is that the P3b reflects context updating: when an incoming stimulus deviates from the representation maintained in working memory, the representation is updated and a P3b is generated (Donchin & Coles, 1988; Donchin, 1981; Polich, 2007; but see Verleger, 1988). Implicit in this and other theories is the idea that the P3b is only evoked once a stimulus has been categorized. According to this stimulus-evaluation account, P3b latency can be used to divide processing into two general stages – stimulus-related and response-related. P3b latency is strongly affected by the duration of stimulus-related processing. Many studies have confirmed that P3b latency increases when stimulus categorization is difficult (Coles, Gratton, Bashore, Eriksen, & Donchin, 1985; Kutas, McCarthy, & Donchin, 1977; Magliero, Bashore, Coles, & Donchin, 1984; Twomey, Murphy, Kelly, & O'Connell, 2015). Alternatively, P3b latency is weakly affected by manipulations that prolong response planning, particularly if there is little temporal overlap between stimulus categorization and motor preparation

* Corresponding author.
E-mail address: mmw188@gmail.com (M.M. Walsh).

http://dx.doi.org/10.1016/j.biopsycho.2016.11.011
0301-0511/

(McCarthy & Donchin, 1981; Smulders, Kok, Kenemans, & Bashore, 1995; Verleger, 1997).

The relationship between the P3b and responding remains controversial, however. Donchin and Coles ascribed the P3b (i.e., the P300) a role in *strategic* processing (1988). They proposed that the P3b is not involved in immediate responding per se, but rather in some overarching process that operates in parallel or afterwards. For example, stimulus categorization could influence future behavior by causing strategy shifts or changes in cognitive control. In support of this view, the P3b sometimes follows the response, suggesting that it does not directly relate to responses (for a review, see Donchin & Coles, 1988). Additionally, correlations between P3b latencies and response times (RTs) are sometimes weak or absent, reinforcing the idea that the P3b does not directly contribute to responses.

Other evidence indicates that the P3b plays a role in *tactical* processing, however (O'Connell et al., 2012; Verleger, Grauhan, & Śmigasiewicz, *in press*; Verleger, Jaśkowski, & Wascher, 2005; for a review, see Nieuwenhuis, Aston-Jones, & Cohen, 2005). By this view, the P3b bridges stimulus categorization and response preparation within trials. In support of this view, many studies have found that the P3b precedes responses, and that P3b latencies correlate with RTs (for a review, see Verleger, 1997). This suggests that responses depend on the mental process reflected in the P3b. Building on these findings, O'Connell et al. (2012) demonstrated that the P3b could be treated as a build-to-decision variable that peaked with behavioral responses in a range of tasks that varied stimulus contrast and modality. Further evidence comes from studies that compare temporal synchronization of the P3b with stimulus onset and response commission. Verleger et al. (2005) compared the amplitudes of the P3b in stimulus-locked and response-locked waveforms. They reasoned that if the P3b relates only to stimulus processing, it would be time-locked to stimulus onset and its amplitude would be largest in stimulus-locked waveforms. Conversely, if the P3b signifies the conclusion of stimulus processing and the organization of a response, its amplitude would be equal in stimulus-locked and response-locked waveforms. P3b amplitude was, in fact, equal in stimulus-locked and response-locked waveforms, consistent with its role in tactical processing. To summarize, the tactical and strategic views both assume a relationship between stimulus categorization and the P3b. They differ, however, in whether they ascribe the P3b a role in immediate responses or future behavior.

Existing studies of the P3b latency are limited in key ways. Some have found little to no relationship between P3b latencies and RTs despite the fact that the experimental manipulations would be expected to affect stimulus processing (for a review, see Verleger, 1997). This may reflect shortcomings of the underlying stimulus-evaluation account and the tactical view of the P3b, or limitations of the methods used to measure P3b latency. Additionally, the relationship between P3b latencies and RTs is typically assessed by comparing their averages in different conditions (McCarthy & Donchin, 1981; Ritter, Simson, Vaughan, & Macht, 1982), or by binning trials according to RTs and comparing average P3b latencies from trials with fast versus slow responses (Friedman, 1984; Roth, Ford, & Kopell, 1978). Both approaches ignore information about trial-by-trial variability in P3b latencies and RTs. The peak latency of the P3b varies within individuals on the order of tens of milliseconds (Nishida, Nakamura, Suwazono, Honda, & Shibasaki, 1997; Pfefferbaum, Ford, Roth, & Kopell, 1980a,b), raising the possibility that trial-by-trial fluctuations in P3b latencies relate to changes in RTs (but see Ramchurn, de Fockert, Mason, Darling, & Bunce, 2014 for an alternate view).

The majority of studies comparing P3b latencies to RTs use tasks where stimuli must be categorized before responding. Simple detection tasks, which require merely detecting the onset of a single repeating target stimulus, also evoke a P3b (Mertens & Polich, 1997; Polich & Heine, 1996; Polich, Eischen, & Collins, 1994). The finding that a continuously repeating target produces a P3b may seem to undermine the idea that the component reflects context updating. However, given the ancillary assumption that the representation of context held in memory decays over time, the re-occurrence of the target may serve to restore the representation (Gonsalvez, Barry, Rushby, & Polich, 2007). Even if such an update is obligatory, it is not needed to respond in simple detection tasks. If the P3b occurs after stimulus categorization is complete, but initial encoding is all that is needed to respond, the P3b may come after responses and be less strongly associated with them in simple detection versus choice RT tasks.

In this paper, we present a novel method for detecting significant events in the EEG signal. These events relate to early visual components, and later cognitive components such as the P3b. We apply the method to data from three paradigms commonly used to study the P3b in order to understand how the component and its relationship to behavioral responses differs among tasks. We describe the P3b task variants and the analytical technique next.

1.1. P3b task variations

The P3b is robustly evoked in three variants of the oddball paradigm. In the *single stimulus* variant, targets are presented after unpredictable temporal delays (Mertens & Polich, 1997; Polich & Heine, 1996; Polich et al., 1994). No other stimuli are shown, and participants are instructed to respond to targets. In the *two stimulus* variant, target and standard distractor stimuli are presented (Duncan-Johnson & Donchin, 1977; Squires, Squires, & Hillyard, 1975). Target stimuli appear infrequently, and participants are instructed to respond to targets. Finally, in the *three stimulus* variant, targets, standard distractors, and non-target distractors are presented (Comerchero & Polich, 1999; Conroy & Polich, 2007). Standard and non-target distractors differ in that non-target distractors are presented at the same (low) frequency as targets. In all three tasks, targets produce a centroparietal positivity consistent with the P3b. Non-target distractors, which appear infrequently in the three-stimulus paradigm, produce a more frontocentral P3a. Finally, standard distractors produce little to no positive modulation during the time window of the P3.

Katayama and Polich (1996) recorded EEG from participants as they completed one-, two-, and three-stimulus oddball paradigms. P3b peak amplitude was equivalent across tasks, while P3b peak latency was shorter in the one-stimulus oddball paradigm as compared to the two- and three-stimulus variants (see also, Walhovd & Fjell, 2001). Within individuals, P3b amplitudes were correlated between all pairs of tasks, as were P3b latencies. Other studies have also found that the topographical distribution of the P3b is consistent across one- and two-stimulus oddball tasks (Croft, Gonsalvez, Gabriel, & Barry, 2003), as are the neural generators inferred using source modeling (Tarkka & Stokic, 1998).

1.2. HSMM-MVPA

Inspired in part by ERP-based mental chronometry pioneered with the P3b (Duncan-Johnson, 1981; Kutas et al., 1977; Meyer, Osman, Irwin, & Yantis, 1988; Pfefferbaum et al., 1980a,b), we have developed a novel method for decomposing the EEG signal into a sequence of processing stages. The method reveals information about the number and latencies of processing stages in a task, along with their associated scalp distributions (Anderson, Zhang, Borst, & Walsh, 2016; Zhang, Walsh, & Anderson, in press). Here, we apply the method to three paradigms known to evoke the P3b to better understand the types and timing of mental operations that occur in each.

The standard approach to analyzing EEG data in the time domain is to segment the continuous recording, align the segments to observable events such as the onset of a stimulus or the commission of a response, and average the signal across the segments to create event-related potentials, or ERPs (Luck, 2014; Vaughan, 1969; Walter, Cooper, Aldridge, McCallum, & Winter, 1964). ERPs contain characteristic sequences of negative- and positive-going deflections, including the P3b. Comparison of the ERPs obtained for different conditions of an experiment can reveal whether, when, and for how long a manipulation impacted mental processing. Although this approach to EEG analysis has been extremely informative, it has two important limitations. First, neural signals that are only weakly locked to observable events are attenuated in average waveforms because they occur at different times during different trials. This is especially problematic for later signals that reflect variability from the durations of processes that precede them. Second, ERPs represent average activity and discard information about differences in when various components appeared within each trial. Trial-by-trial variability in component latencies could be informative with respect to understanding variability in overt behavioral responses.

To overcome these limitations, we developed a technique for applying hidden semi-Markov models and multivariate pattern analysis to EEG data (HSMM-MVPA; Anderson et al., 2016; Zhang et al., in press). The method assumes that in the absence of task-related processing, the background EEG can be modeled as ongoing sinusoidal noise (Fig. 1, left), the amplitude of which varies by frequency according to a known power spectrum (Yeung, Bogacz, Holroyd, Nieuwenhuis, & Cohen, 2007). Stimulus onset initiates task-related processing, and transitions from one cognitive stage to the next are marked by the superposition of *bumps* on the background EEG. A bump is a multidimensional distribution of scalp activity, and is modeled as a half cycle of a 10 Hz sine wave (Fig. 1, middle). The bump rises, peaks, and falls over 50 ms. The recorded EEG contains ongoing sinusoidal noise along with the signal from one or more bumps (Fig. 1, right). This conceptual model is consistent with the classical theory of ERP generation (Yeung et al., 2007), in which ERPs arises from the addition of event-locked signals to the background EEG. Others have argued that ERPs arise when an event causes the phase of an oscillation in a certain frequency band to be reset, causing its peaks and troughs to be aligned across trials (Basar, 1980). Although these theories have different conceptual bases, both predict the appearance of phasic bumps time-locked to experiment events (Yeung, Bogacz, Holroyd, & Cohen, 2004, 2007)

The intervals between bumps, referred to as *flats*, contain only sinusoidal noise with mean zero. The durations of flats differ on a trial-by-trial basis, reflecting variability in the latencies of the underlying mental processes. Flat durations are modeled as random variables drawn from gamma distributions (Fig. 2). The HSMM estimates when bumps occurred during each trial, and the MVPA estimates the multidimensional scalp activity associated with a bump. Using HSMM-MVPA, it is possible to infer the number, timing, and topography of bumps that maximize the likelihood of the EEG data.

We previously applied HSMM-MVPA to two studies of associative recognition and a study using the Sternberg Task (Anderson et al., 2016; Zhang et al., in press). In all of those cases, a computational cognitive model for the task had already been created using the integrated cognitive architecture ACT-R (Anderson, 2007). The models contained predefined numbers of processing stages, and specified how the durations of stages were impacted by the various experimental manipulations. The main contribution of HSMM-MVPA was to test whether the numbers and durations of stages inferred from the EEG were consistent with the computational models. This constitutes a top-down application of HSMM-MVPA for testing existing theories. HSMM-MVPA can also be used in a bottom-up manner to infer the number and durations of processing stages directly from the EEG. We adopt the bottom-up approach in this paper.

1.3. Overview of current experiments

We recorded EEG from participants as they performed three tasks: the Psychomotor Vigilance Task (PVT; Dinges & Powell, 1985), the Hitchcock Radar Task (Hitchcock, Dember, Warm, Moroney, & See, 1999), and the 3-Stimulus Task (Comerchero & Polich, 1999). In the PVT, participants monitored the screen for the onset of a millisecond counter. They responded when the counter appeared. In the Hitchcock Radar Task, participants saw frequent non-critical signals and infrequent critical signals. They responded to critical signals only. Finally, in the 3-Stimulus Task, participants saw frequent distractors, rare distractors, and targets. They responded to targets only. The particular tasks were selected for reasons unrelated to the current paper. Yet they exemplify the one-, two-, and three-stimulus oddball paradigms commonly used to study the P3b.

We applied HSMM-MVPA to EEG data from these tasks to address two questions. First, does HSMM-MVPA reveal a stage with a bump corresponding to the P3b? Based on the existing literature, we expected to see such a bump after counter onset in the PVT, after critical signals in the Radar Task, and after targets in the 3-Stimulus Task. Second, is the latency of the P3b stage associated with RTs? Both the Radar and 3-Stimulus Task require categorizing stimuli before responding. As such, we expected responses to follow, and be synchronized with, the P3b stage in trials of those tasks. The PVT only requires detecting a stimulus before responding. Consequently, it was unclear whether responses would precede or follow the P3b in the PVT, and whether RTs would relate to P3b latencies.

2. Methods

2.1. Participants

Thirty individuals from the University of Dayton community participated in a single two hour session for monetary compensation (10 men and 20 women, ages ranging from 18 to 36 with a mean of 22.7). 27 were right-handed. None reported a history of neurological impairment.

2.2. Tasks

Each participant completed three tasks: the Psychomotor Vigilance Task (PVT), the 3-Stimulus Task, and the Radar Task. The tasks were separated by five-minute breaks. Because the Radar Task was longest, all participants completed that task first, followed by the PVT and the 3-Stimulus Task. We describe the tasks, and later the results, in order of task difficulty as inferred from RTs. Responses were fastest in PVT, followed by the 3-Stimulus Task, and the Radar Task.

2.2.1. Psychomotor vigilance task

The PVT is frequently used in sleep research (Dinges & Powell, 1985; Lim & Dinges, 2008). Participants monitored a display for the onset of a visual millisecond counter within a box that was 5 cm wide by 1 cm high. The counter started from zero and continuously incremented until the participant responded, or until 30 s passed. Elapsed time remained on the screen for 1 s after the participant responded and served as performance feedback. The next trial began from 1 to 9 s after feedback was extinguished. The cumulative duration of the inter-trial interval (including 1 s for feedback) was randomly drawn from a uniform integer distribution ranging

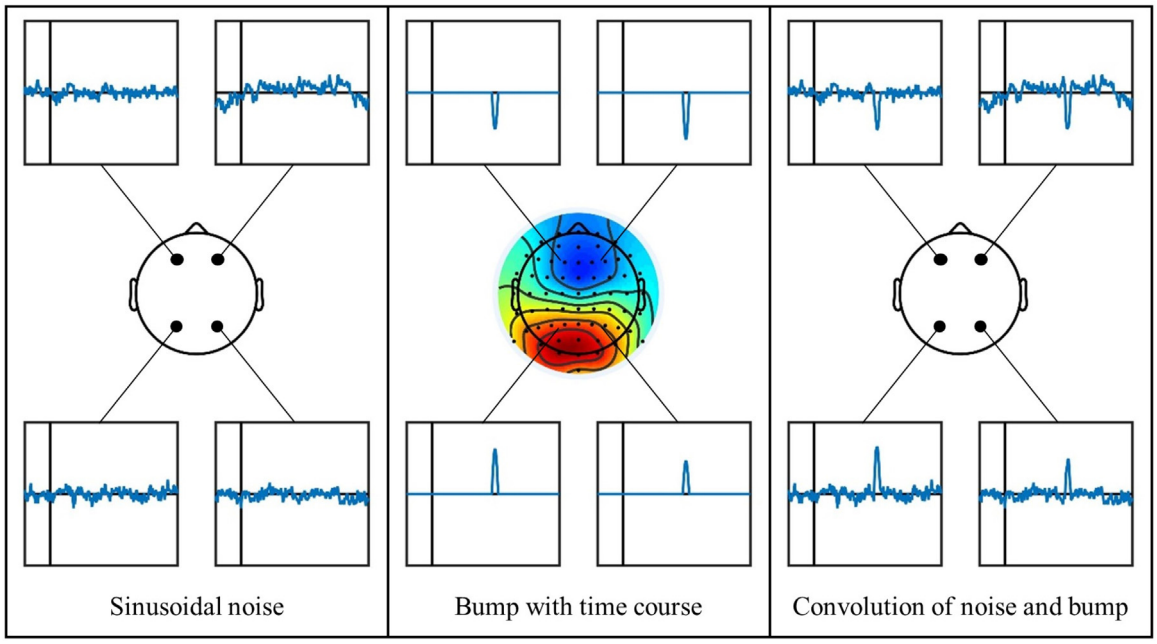

Fig. 1. Conceptual model of ERP generation, and assumptions underlying bump analysis.

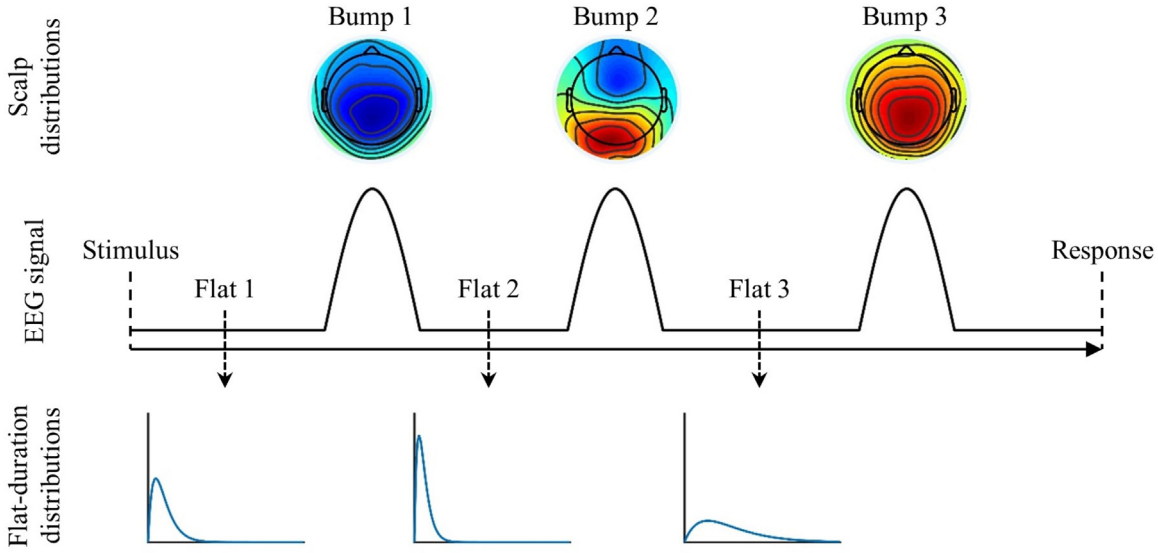

Fig. 2. Examples of three multidimensional bumps along with flat-duration distributions.

from 2 to 10 s. The counter and feedback appeared in white against a black background.

Participants were instructed to respond as quickly as possible, but to avoid responding before the stimulus appeared. They pressed the "J" key on a standard keyboard using their right index finger. Participants were given 1 min of practice on the PVT. Following a short break, they performed the task continuously for 10 min.

2.2.2. 3-Stimulus task

The 3-Stimulus Task was based on the visual condition of two prior studies (Comerchero & Polich, 1999; Conroy & Polich, 2007). In each trial, participants viewed a target, a non-target distractor, or a standard distractor. Targets and non-target distractors each appeared in 10% of trials, and standard distractors appeared in the remaining 80% of trials. The target was a large circle (2.5 cm diameter), the non-target distractor was a large square (2.5 cm height), and the standard distractor was a small circle (1.6 cm diameter). Stimuli were presented in blue against a black background. A stimulus appeared once every 2 s, and remained on the screen for 75 ms.

Participants were instructed to respond to targets and to ignore non-target and standard distractors. They pressed the "J" key on a standard keyboard using their right index finger. Participants were given 20 practice trials. Following a short break, they completed 4 blocks of 90 trials. After each block, they received feedback about the percentages of targets detected (i.e., *hits*), and non-target and standard distractors ignored (i.e., *correct rejects*).

2.2.3. Radar task

The Radar Task was based on a study by Hitchcock et al. (1999). A stimulus display with a filled central circle (1.6 cm diameter) and three concentric rings (2.8 cm, 3.9 cm, and 5.1 cm diameters) appeared in each trial (Fig. 3). Two line segments (0.15 cm wide by 1.4 cm long) were superimposed on the concentric rings. Participants were told that the filled circle nominally corresponds to

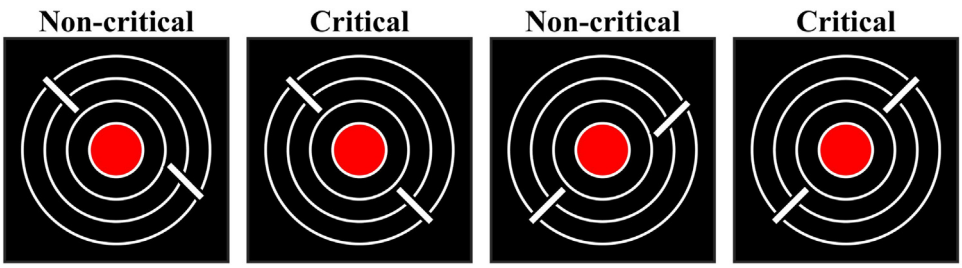

Fig. 3. Example configurations from Radar Task.

a city, and the line segments to two jets. They were instructed to respond if the two jets were on a collision path with one another (i.e., the line segments were colinear). Critical events occurred in 3.3% (10 out of every 300) trials. Stimulus displays were presented in white against a black background. A new display appeared every 2 s and remained on the screen for 300 ms.

Participants were instructed to respond to critical events and to ignore non-critical events. They pressed the "J" key on a standard keyboard using their right index finger. Participants were given 200 practice trials. During practice, they received feedback about the percentages of critical events detected and non-critical events ignored after every 50 trials. Following a short break, they performed the task continuously for 40 min (1600 trials) without feedback or breaks.

2.3. EEG recording and analysis

Stimuli were presented on a 60 Hz LCD monitor placed approximately 60 cm from participants. The EEG was recorded from 64 Ag-AgCl sintered electrodes (10–20 system) using a Biosemi Active II System (BioSemi, Amsterdam, Netherlands). Voltage offsets were reduced to 40 mV or less, and the resting EEG was checked for any problematic electrodes prior to recording.[1] The EEG was re-referenced online to the combined common mode sense (CMS) and driven right leg (DRL) circuit. Electrodes were also placed on the right and left mastoids. Scalp recordings were algebraically re-referenced offline to the average of the right and left mastoids. The EEG signals were filtered with a bandpass of 0.1–70.0 Hz and were digitized at 512 Hz.

The EEG recording was decomposed into independent components using the EEGLAB FastICA algorithm (Delorme & Makeig, 2004). Components associated with eye blinks were automatically identified and projected out of the EEG recording. Epochs of 1200 ms (including a 200 ms baseline) were then extracted from the continuous recording and corrected over the prestimulus interval. Epochs containing voltages above +100 μV or below −100 μV were excluded from further analysis. One participant with less than 30% artifact free data was excluded from the PVT, two participants were excluded from the 3-Stimulus Task, and two participants were excluded from the Radar Task.

The P3b is typically defined in terms of its topographical distribution along the midline and over the frontal, central, and parietal scalp (Comerchero & Polich, 1999; Conroy & Polich, 2007). Accordingly, we averaged across contiguous electrodes to create three regions: a frontal region (AF3, Fpz, AFz, Fz, AF4), a central region (C1, FCz, Cz, CPz, C2), and a parietal region (P1, CPz, Pz, POz, P2). P3b latency varies by task. For the PVT, we analyzed waveforms evoked by the onset of the visual millisecond counter. We quantified the P3b as mean voltage from 250 to 500 ms (Witkowski et al., 2015). For the 3-Stimulus Task, we compared waveforms evoked by the onset of targets, non-target distractors, and standard distractors. We quantified the P3b as mean voltage from 350 to 500 ms (Comerchero & Polich, 1999). Lastly, for the Radar Task, we analyzed waveforms evoked by the onset of critical and non-critical events. Only one previous EEG study of the Radar Task has been published (Kamzanova, Kustubayeva, & Matthews, 2014), and that study did not analyze ERPs. As such, the interval used to quantify the P3b (450–800 ms) in the Radar Task was based on visual inspection of the waveforms. For all analyses involving factors with more than two levels, we adjusted the *p* values using the Greenhouse-Geisser correction.

2.4. HSMM-MVPA analysis

HSMM-MVPA models the variability of endogenous ERP components that may otherwise be distorted or lost in average waveforms (Anderson et al., 2016; Zhang et al., *in press*). The method identifies brief periods of activity characterized by distinctive patterns of voltages across the scalp called *bumps*. Bumps occur at variable latencies in the single-trial EEG, and signify changes in information processing. Bumps are separated by *flat* periods where the mean of the ongoing sinusoidal noise is zero. Flat durations vary by trial, and are treated as random variables drawn from gamma distributions. The HSMM estimates the locations of bumps during each trial, and the MVPA estimates their scalp topographies.

The EEG data were pre-processed in the same manner as for the ERP analyses. Two additional steps of dimensionality reduction were performed to increase the efficiency and tractability of HSMM-MVPA. First, the data were interpolated and down-sampled to 100 Hz (i.e., 10-ms samples). Second, because signals from different EEG electrodes are highly inter-correlated, spatial PCA was performed and scores of the first 10 PCA components were retained. These accounted for 86% of variance in the EEG signal during the PVT, 85% of variance during the 3-Stimulus Task, and 83% variance during the Radar Task.[2] The PCA components were z-scored for each trial. The final data consisted of 10 orthogonal PCA components sampled every 10 ms and with constant mean and variability across trials.

HSMM-MVPA contains three ancillary assumptions. First, bumps have a 50 ms duration (i.e., 5 samples) and a half-sine shape. The use of such narrow bumps promotes precision in the identification of stage boundaries, and allows HSMM-MVPA to simultaneously identify components with power focused in higher frequencies (e.g., the N1) and lower frequencies (e.g., the P3b). Previously, we found that HSMM-MVPA with a 50 ms bump accurately

[1] Electrode offset is the voltage measured between CMS and each active electrode, and reflects in part the quality of the electrode/gel/skin contact. We achieved offsets far less than the maximum recommended value of 40 mV (BioSemi, Amsterdam, Netherlands).

[2] We used PCA for noise reduction and data compression, rather than to isolate latent components. As such, we retained unrotated scores to maximize the variance accounted for. For comparison, the scores of the first 10 components returned by Promax with a kappa value of three (Dien, 2010) accounted for about 20% less variance in the PVT (69%), the 3-Stimulus Task (64%), and the Radar Task (65%).

recovered the temporal location of 3 Hz signals (Anderson et al., 2016), which is important given that the P3b reflects power in the delta and theta bands (Demiralp, Ademoglu, Comerchero, & Polich, 2001). Second, bumps occur serially and do not overlap. Third, flat durations are modeled as a gamma distribution with a fixed shape parameter 2 and a scale parameter estimated to fit the data. See Anderson et al. (2016) for a detailed discussion of these assumptions, and demonstrations of the robustness of the method against violations of each.

In our previous studies, we used HSMM-MVPA to place *n* bumps between stimulus onset and response commission (Anderson et al., 2016; Zhang et al., *in press*). Fitting an n-bump HSMM requires estimating *n + 1* stage distributions to describe the *n + 1* periods of time collectively demarked by stimulus onset, the *n* bumps, and the response. Most trials in the current tasks do not end with responses. As such, we fit n-bump HSMMs to the EEG from the first full second of every trial in all tasks.[3] This requires estimating *n* stage distributions to describe the duration of the flat preceding the first bump, and the durations of the flats separating the subsequent bumps. Since the specific interval of time used in the HSMM-MVPA analysis did not end with a response and has no special significance, we did not estimate a flat duration from the last bump to the end of the interval.

Estimating an n-bump HSMM also requires estimating the topographical distributions of the *n* multidimensional bumps. These are intended to capture signals, present in the 10 PCA dimensions, during the 50 ms windows where bumps are centered. A bump is defined as a 2-dimensional matrix of values B_{ij} where *j* indexes the 5 samples that make up the 50 ms bump and *i* indexes the 10 PCA dimensions. The bump values are calculated as $B_{ij} = P_j{}^*M_i$ where P_j are the 5 sample weights (i.e., 0.309, 0.809, 1.000, 0.809, and 0.309) that describe the shape of a half-sine and M_i are the 10 PCA magnitudes for bump *i*. The values of P_j are fixed based on the half-sine shape, and the values of M_i are estimated for each bump. To summarize, finding *n* bumps requires estimating *n* scale parameters for the gamma distributions that characterize the durations of the flats that precede each bump, and 10 x *n* magnitudes M_i to capture the observed PCA scores during the 50 ms windows that each bump spans.

The likelihood of a set of flat durations $(t_1, t_2, \ldots t_n)$ during a trial is given by

$$P(t_1, t_2 \ldots t_n) = \left(\prod_{k=1}^{n} g_2(t_k, b_k)\right)\left(\prod_{d=1}^{10} P(PCA_d)\right).$$

The first inner product in the equation describes the likelihoods of the flat durations based on the gamma distributions they are drawn from. HSMM-MVPA estimates the scale parameter (b_k) for each gamma distribution. The second inner product in the equation describes the likelihoods of the scores along each of the 10 PCA dimensions. These are derived from the sum of the squared deviations between the observed PCA scores and the model scores, which are zero during the flats and non-zero during the bumps (Anderson et al., 2016). The probability of the data in a trial is the sum of the probabilities of all possible placements of bumps within the trial, each calculated as above.

The HSMM-MVPA estimates parameters (t_n stage durations and M_n bump topographies) that maximize the likelihood of the data over all trials of the experiment. The HSMM uses an iterative expectation maximization algorithm (Dempster, Laird, & Rubin, 1977). The dynamic programming techniques associated with the HSMM make tractable the computation of probabilities of the very large number of ways of partitioning each trial into bumps and flats (Yu, 2010). When the iterative algorithm converges on a set of parameters than maximize the probability of the data, the algorithm also produces estimates of the probabilities that each bump occurred at each sample of each trial.

2.5. Comparison of neural processing to RTs

HSMM-MVPA returns a sequence of bumps that appear in the EEG for each trial. The bumps come after flats, whose durations vary across trials according to gamma distributions with scale parameters estimated by the HSMM. Because the flat durations vary across trials, they can be used to create a predictive model of RTs. We used R (R Core Team, 2016) and *lme4* (Bates, Maechler, Bolker, & Walker, 2015) to perform a linear mixed effects analysis of the relationship between the trial-level bump latencies and RTs (Saville et al., 2015). If the cognitive processes reflected in a bump are needed to respond, we expected that inclusion of the time preceding that bump (i.e., the duration of the flat) would improve the predictive model of RTs.

We compared the predictive model based on HSMM-MVPA with two alternate techniques commonly used to quantify P3b latency: peak picking and template matching (Woody, 1967). Following Jaśkowski and Verleger (2000), we applied a 3.5 Hz low-pass filter to the single-trial data. In the case of peak picking, we quantified P3b latency as the moment of peak amplitude at Pz. In the case of template matching, we generated a positive half-sinusoid of 300-ms length. We calculated the cross-covariance between the template and the signal during 300 ms time windows across every trial, and quantified the P3b latency as the center of the window that maximally covaried with the template. We then computed a new template using the average of the observed data from the identified windows, and iterated on the procedure until the template converged on a stable average. For both peak picking and template matching, we used linear mixed effects models to examine the relationship between P3b latencies and RTs.

3. Behavioral results

3.1. Task data

Responses in the PVT are often divided into three categories: *false starts* occur before or within 150 ms of counter onset; *alert responses* occur from 150 ms to 500 ms after counter onset; and *lapses* occur more than 500 ms after counter onset. Table 1 shows the distribution of response types along with the mean duration of alert responses. Alert RTs and number of lapses were comparable to PVT results obtained elsewhere with well-rested participants (for reviews, see Basner & Dinges, 2011; Lim & Dinges, 2008).

In the 3-Stimulus Task, participants were instructed to respond to targets, and to ignore non-target and standard distractors. Table 1 shows the percentage of hits and false alarms, along with the mean duration of hits. Although false alarms were infrequent, participants reliably responded to more non-target than standard distractors, $t(27) = 3.79$, $p < 0.001$. Mean RTs, hit rates, and the relative frequencies of false alarms were consistent with outcomes from other studies of the 3-Stimulus Task that used visual stimuli and were of comparable difficulty (Comerchero & Polich, 1999; Demiralp et al., 2001).

In the Radar Task, participants were instructed to respond to critical events and to ignore non-critical events. Table 1 shows the percentage of hits and false alarms, along with the mean duration of hits averaged across the entire experiment. Mean hit rates across the 40 min session were comparable to those previously reported (Hitchcock et al., 1999).

[3] Responses occurred within 1 s in more than 99% of trials from the PVT and the 3-Stimulus Task, and in more than 98% trials from the Radar Task

Table 1
Mean response time and distribution of response types with mean standard error in parenthesis.

Psychomotor Vigilance Task		3-Stimulus Task		Radar Task	
Measure	Value	Measure	Value	Measure	Value
Response time (ms)	340 (8)	Response time (ms)	517 (10)	Response time (ms)	730 (18)
False starts (%)	0.7 (0.2)	Target hits (%)	97.5 (0.7)	Target hits (%)	85.1 (2.3)
Alert responses (%)	95.0 (1.2)	False alarms (%)		False alarms (%)	4.5 (1.2)
Lapses (%)	4.3 (1.2)	Non-target	1.9 (0.5)		
		Standard	0.1 (0.1)		

3.2. Between-task comparisons

We restricted analyses to 26 participants with complete data from all tasks. Responses were faster in the PVT than in the 3-Stimulus Task ($t(25) = 19.2$, $p < 0.001$) and the Radar Task ($t(25) = 21.2$, $p < 0.001$), and responses were faster in the 3-Stimulus Task than in the Radar Task ($t(25) = 11.0$, $p < 0.001$). Hit rate was higher in the 3-Stimulus Task than in the Radar Task ($t(25) = 5.24$, $p < 0.001$), and false alarm rate was lower ($t(25) = 2.77$, $p < 0.05$). Mean RTs from the 26 participants were moderately correlated pairwise between the tasks (ranging from $r = 0.18$ to 0.53), as were hit rates and false alarms between the 3-Stimulus and Radar Tasks ($r = 0.30$ and 0.10, respectively).

4. ERP results

4.1. Task data

Fig. 4 shows waveforms from the PVT time-locked to the onset of the visual counter. Mean amplitude of the waveforms from 250 to 500 ms varied by region, $F(2,56) = 45.6$, $p < 0.001$. Voltages were most positive over the parietal region (Table 2). For each participant, we treated peak latency of the P3b as the time of the maximum voltage in the averaged waveform over the parietal region from 250 to 800 ms (the complete range of the time windows used across the three tasks). On average, the P3b peaked at 429 ± 25 ms.

Fig. 4 also shows waveforms from the 3-Stimulus Task time-locked to the onset of targets, non-target distractors, and standard distractors. A 3 (region) by 3 (stimulus type) repeated measures ANOVA on mean amplitude from 350 to 500 ms revealed main effects of region, $F(2, 54) = 41.6$, $p < 0.001$, and stimulus type, $F(2, 54) = 24.5$, $p < 0.001$. The interaction between region and stimulus type was also significant, $F(4, 108) = 38.7$, $p < 0.001$. The P3 to targets (i.e., the P3b) had a parietal maximum, whereas the P3 to non-target distractors (i.e., the P3a) was more centrally distributed (Table 2). Standard distractors did not produce a P3. The P3b during target trials peaked at 458 ± 16 ms.

Finally, Fig. 4 shows waveforms from the Radar Task time-locked to the onset of critical and non-critical events. A 3 (region) by 2 (stimulus type) repeated measures ANOVA on mean amplitude from 450 to 800 ms revealed main effects of region, $F(2, 54) = 8.329$, $p < 0.001$, and stimulus type, $F(1, 27) = 17.8$, $p < 0.001$. The interaction between region and stimulus type was also significant, $F(2, 54) = 38.6$, $p < 0.001$. The P3 had a parietal maximum and was greater for critical versus non-critical events (Table 2). The P3b to critical events peaked at 611 ± 30 ms.

4.2. Between-task comparison

We compared the amplitudes of the P3b evoked by the visual millisecond counter in the PVT, targets in the 3-Stimulus Task, and critical events in the Radar Task for the 26 participants with complete data. P3b amplitude over the parietal area was smaller in the Radar Task than in the PVT ($t(25) = 3.82$, $p < 0.001$) and the 3-Stimulus Task ($t(25) = 6.82$, $p < 0.001$), and was similar between the PVT and the 3-Stimulus Task ($t(25) = 1.81$, $p < 0.1$). We also compared the peak latencies of the P3b. Peak latency occurred later in the Radar Task than in the PVT ($t(25) = 4.64$, $p < 0.001$) and the 3-Stimulus Task ($t[25] = 4.60$, $p < 0.001$), and was similar between the PVT and the 3-Stimlus Task ($t(25) = 1.09$, $p > 0.2$). The peak latency trends are only partially consistent with the behavioral results – participants responded slower in the 3-Stimulus Task than in the PVT, yet P3b peak latencies did not differ between the two tasks. P3b peak amplitudes from the 26 participants were moderately correlated pairwise between tasks (PVT and 3-Stimulus Task: $r = 0.49$; PVT and Radar Task: $r = 0.23$; 3-Stimulus Task and Radar Task: $r = 0.21$), whereas peak latencies were not (all $r < 0.05$).

5. HSMM-MVPA results

5.1. Psychomotor vigilance task

We compared how well two- and three-bump HSMMs accounted for the EEG data from the PVT. We fit each HSMM to *n-1* participants' data and used the parameterized model to predict the remaining participant's data (leave-one-out cross validation, or LOOCV). Adding more bumps involves estimating more parameters. As such, log-likelihoods based on fit always increase from two to three bumps, but log-likelihoods based on LOOCV only increase if the additional bump captures structure in the data that generalizes to other participants. Table 3 shows the average log-likelihood values per participant and trial based on LOOCV. The addition of a third bump improved out-of-sample prediction for most participants ($p < 0.05$ by a two-tailed sign test).

Fig. 5 shows the bump topographies along with their mean latencies from the PVT (upper-left panel). The third bump corresponds to the P3b with its posterior distribution and late onset. The second bump resembles the N2b, an anterior negativity that precedes the P3 and is sensitive to many of the same experimental factors (Folstein & Van Petten, 2008). Activity related to the second and third bumps is visible in the ERP waveforms at the corresponding latencies (Fig. 4).

The first bump had a short latency and negative polarity over the parietal scalp, yet the ERPs contain no evidence of such an early deflection. This dissociation relates to trial-by-trial variability in the latency of the first bump. Uncertainty about the latency of a bump within a trial is represented by a probability distribution with values corresponding to the likelihood that the bump is centered at each sample. The latency of a bump in a trial is calculated as the sum of the time points multiplied by the probability that the bump is centered at each sample – that is, a weighted average of the time points.

We divided trials into two bins according to the estimated latency of the first bump using a median split. Trials in both bins had an early negative-going deflection followed by a positive peak (Fig. 6, left). Further, the waveforms appeared to contain an oscillation that originated during the prestimulus interval and extended through the first bump. We fit sine waves to averaged data from the two sets of trials from −200 ms to the time where voltages exceeded zero following the first bump (90 ms for trials where

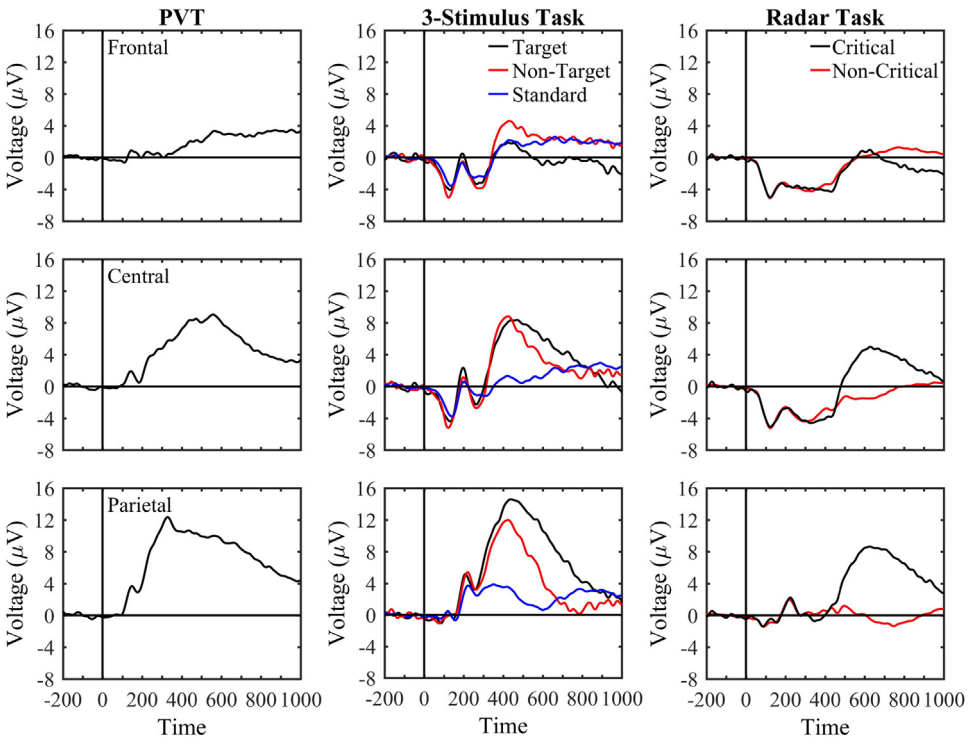

Fig. 4. Stimulus-locked waveforms from three tasks, and from electrode clusters centered over frontal, central, and parietal regions.

Table 2
Mean voltages over frontal, central, and parietal regions in three tasks with mean standard error in parenthesis.

	PVT	3-Stimulus Task			Radar Task	
Region		Target	Non-Target	Standard	Critical	Non-Critical
Frontal	0.97 (0.95)	1.30 (1.14)	3.52 (1.11)	1.46 (0.80)	−0.11 (1.09)	0.10 (0.61)
Central	6.01 (1.25)	6.90 (1.28)	7.15 (1.23)	0.77 (0.61)	2.99 (1.08)	−0.92 (0.42)
Parietal	9.92 (1.31)	12.54 (1.22)	10.12 (1.15)	2.94 (0.59)	6.01 (1.02)	−0.16 (0.50)

Table 3
Log likelihood per trial and participant based on LOOCV from 2- and 3-bump HSMMs, and number of participants better predicted by a 3-bump HSMM.

	2 Bumps	3 Bumps	Number Participants 2 Bumps < 3 Bumps
PVT			
Counter	1.04	1.88	25/29
3-Stimulus Task			
Target	1.25	1.96	21/28
Non-target distractor	0.87	1.30	21/28
Standard distractor	0.49	0.51	17/28
Radar Task			
Non-critical	0.67	0.85	19/28
Critical	1.28	1.79	20/28

the bump occurred early and 150 ms for trials where the bump occurred late). The amplitudes (1.00 μV versus 1.17 μV) and frequencies (9.6 Hz versus 9.2 Hz) of the fitted sine waves were nearly identical, but the phase of the sine wave for trials where the bump occurred late was shifted to the right by 56 ms. Consequently, the positive and negative cycles of the two groups of waveforms almost perfectly canceled out in the average data. The separation of trials into two bins is arbitrary; the phase of the oscillation varied continuously between trials, and because the oscillation was not stimulus locked, the positive and negative cycles averaged to zero across trials. Makeig et al. (2002) proposed that the N1 arises from alpha ringing (8–12 Hz) over the parietal scalp during visual attention tasks, a point we return to in the discussion.

5.2. 3-Stimulus task

The 3-Stimulus Task included a mixture of trial types that could contain different numbers of bumps; for example, we expected bumps corresponding to the P3 for targets and non-target distractors, but not for standard distractors. Trial types could also contain bumps with different topographies; for example, we expected a more posterior bump for targets (P3b), and a more central bump for non-target distractors (P3a). To test for these differences, we fit 2- and 3-bump HSMMs separately to data from each condition. We constrained the latencies of the first and second bumps to be shared across conditions, and allowed the latencies of the third bump and the topographies of all bumps to vary. Table 3 shows the average log-likelihood values per trial and participant based on LOOCV. The

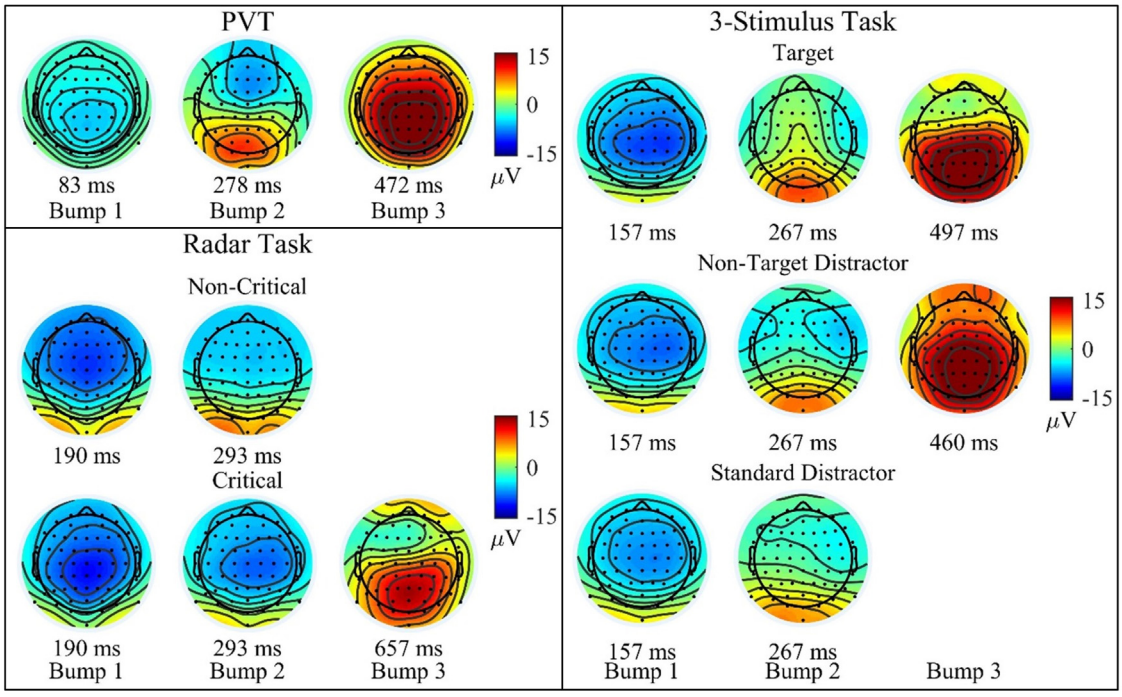

Fig. 5. Topographies and latencies of bumps in the HSMMs from the three tasks.

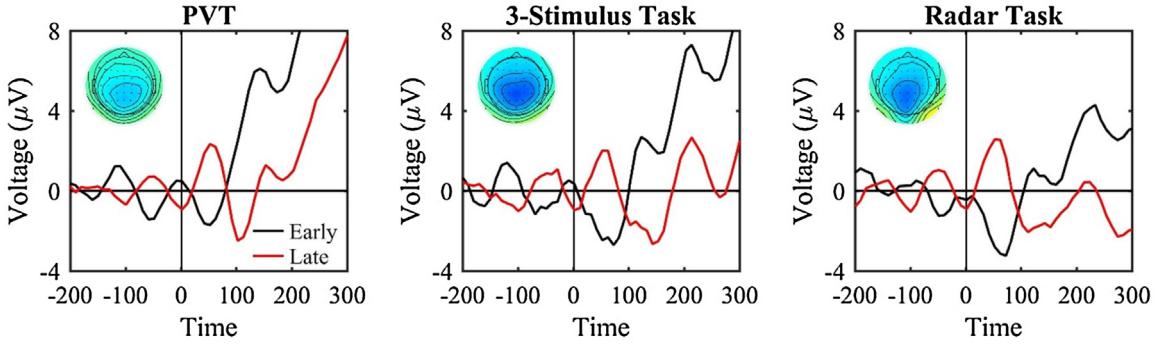

Fig. 6. Waveforms over parietal region binned according to latency of first bump, along with topography of bump across all trials.

addition of a third bump improved out of sample prediction for a majority of participants in target trials and non-target distractor trials ($p < 0.05$ by a two-tailed sign test), but not in standard distractor trials ($p > 0.3$ by a two-tailed sign test).

Fig. 5 shows the 3-bump HSMMs for targets and non-target distractors, and the 2-bump HSMM for standard distractors. We calculated the voltages over the frontal, central, and parietal regions at the peak of the third bump for targets and non-target distractors. The bump had a more posterior distribution for targets than for non-target distractors, reflected in a significant condition-by-location interaction, $F(2, 54) = 40.7$, $p < 0.0001$. This is consistent with the centrally-distributed P3a for non-target distractors and the parietal P3b for targets in the ERP waveforms (Fig. 4).

The topographical distributions of the second and third bumps in target trials of the 3-Stimulus Task closely resembled the corresponding bumps in the PVT (Table 4). This, paired with their consistent latencies caused us to interpret them as the N2/P3 complex in the 3-Stimulus Task as well. Likewise, the topographical distribution of the first bump in the 3-Stimulus task closely resembled the corresponding bump in the PVT (Table 4). Its time course and distribution are consistent with the posterior N1, a component evoked by the presentation of a visual stimulus at an attended location during a task that requires discrimination judgments (Hopf, Vogel, Woodman, Heinze, & Luck, 2002; Makeig et al., 2002; Vogel & Luck, 2000).

Table 4
Mean-squared error (MSE) between voltages in PVT and 3-Stimulus Task across electrodes at the peak latencies of bumps.

	PVT		
3-Stimulus	Bump 1	Bump 2	Bump 3
Bump 1	10.3	59.9	367.9
Bump 2	34.2	13.8	137.1
Bump 3	220.3	101.2	27.4

Although the first bump in the PVT closely resembled the first bump in the 3-Stimulus Task (Table 4), it came about 100 ms earlier in the PVT. If the first bump in the PVT truly reflected an alpha oscillation, as suggested by Fig. 6, then a brain state with a very similar topographic distribution would come 100 ms later. Relatedly, if the first bump in the 3-Stimulus Task reflected a continuation of this oscillation, a brain state with a very similar topographic distribution would come 100 ms earlier.

Table 5
Mean-squared error (MSE) between voltages in the Radar Task, and in the 3-Stimulus Task and PVT across electrodes at the peak latencies of bumps.

	3-Stimulus			PVT		
Radar	Bump 1	Bump 2	Bump 3	Bump 1	Bump 2	Bump 3
Bump 1	1.8	77.9	320.5	15.0	72.3	400.0
Bump 2	1.1	48.7	262.1	7.2	46.2	335.1
Bump 3	170.8	36.6	18.7	115.5	40.5	55.0

To examine this possibility, we fit another HSMM to target trials from the 3-Stimulus Task while constraining the first bump to have a mean latency of 80 ms. The timing and topography of the second and third bumps in the constrained HSMM matched bumps from the original HSMM. The topography of the first bump in the constrained model was nearly identical to the first bump in the PVT ($r = 0.98$, MSE = 4.1). We binned target trials in the 3-Stimulus task according to the estimated latencies of the constrained first bump and observed an oscillation extending from −200 ms to just beyond the bump (Fig. 6, middle). The amplitudes (1.08 μV versus 1.04 μV) and frequencies (8.7 Hz versus 8.0 Hz) of fitted sine waves to the data were similar to one another, but the phase of the sine wave for trials with later bump onset was shifted to the right by 50 ms. Consequently, as in the PVT, the positive and negative cycles of the oscillation averaged to zero across trials. Note, HSMM-MVPA did initially detect a bump at 80 ms in the 3-Stimulus Task. The oscillation had larger amplitude and/or phase synchronization at 150 ms, causing HSMM-MVPA to place the first bump at that time instead.

5.3. Radar task

The Radar Task contained a mixture of trial types that could differ in terms of number and topography of bumps. To examine these differences, we fit 2- and 3-bump HSMMs to data from critical and non-critical trials. We constrained the latencies of the first and second bumps to be the same across conditions, and allowed the latencies of the third bump and the topographies of all bumps to vary. Table 3 shows the average log-likelihood values per trial and participant based on LOOCV. The addition of a third bump improved out of sample prediction for a majority of participants in critical trials only ($p < 0.05$ by a two-tailed sign test).[4]

Fig. 5 shows the 3-bump HSMM for critical trials, and the 2-bump HSMM for non-critical trials. The first and third bumps in critical trials of the Radar Task closely resembled the corresponding bumps in target trials of the 3-Stimulus Task (Table 5). This paired with their similar latencies caused us to interpret them as the N1 and P3b in the Radar Task as well. The second bump in the Radar Task was most similar to the first bump in the 3-Stimulus Task (Table 5). The apparent difference relates to the fact that the P3b appeared earlier in the 3-Stimulus Task, causing it to overlap with the second bump there but not in the Radar Task. This difference aside, we interpret the second bump in the Radar Task as the N2 as well.

To determine whether a posterior oscillation beginning during the prestimulus interval was present in the Radar Task, as in the PVT and 3-Stimulus Task, we fit another HSMM to critical trials that constrained the first bump to have a mean latency of 80 ms. The topography of the first bump in the constrained model was very similar to the first bump in the PVT ($r = 0.90$, MSE = 1.7). We binned critical event trials in the Radar Task according to the estimated latencies of the constrained first bump and observed an oscillation extending from −200 ms to just beyond the bump (Fig. 6, right).

[4] The addition of a third bump improved out of sample prediction in non-critical trials for many participants as well (19/28, $p < 0.1$ by a two-tailed sign test).

Table 6
Negative log likelihood of linear mixed effects models with different fixed factors.

	PVT	3-Stimulus Task	Radar Task
Null Model	14531	5549	12765
HSMM-MVPA			
Bump 1	14426	5460	12717
Bumps 1 and 2	14325	5451	12712
Bumps 1, 2, and 3	14325	5340	12560
P3b peak latency	14514	5502	12764
P3b template matching	14500	5511	12704

The amplitudes (1.00 μV versus 0.90 μV) and frequencies (10.0 Hz versus 8.8 Hz) of fitted sine waves to the data were similar to one another, but the phase of the sine wave for trials with later bump onset was shifted to the right by 30 ms. Consequently, as in the other two tasks, the positive and negative phases of the oscillation averaged to zero across trials.

5.4. Comparison of neural processing to response latencies

The latencies of the bumps vary on a trial-by-trial basis. To determine which bumps relate most to RTs in the three tasks, we compared RTs and bump latencies from 2577 trials in the PVT, 926 target trials in the 3-Stimulus Task, and 2124 critical event trials in the Radar Task (Fig. 7). In the PVT, RTs were most strongly correlated with the latencies of Bump 2, which peaked 39 ms (±6 ms) before responses on average.[5] RTs also correlated with the latencies of Bumps 1 and 3. In the case of Bump 3, the correlation arose because Bumps 1 and 2 impacted the timing of responses as well as cumulative time until Bump 3. In contrast to the PVT, RTs were most strongly correlated with the latencies of Bump 3 in the 3-Stimulus Task and the Radar Task (Fig. 7).[6] Bump 3 peaked 11 ms (±8 ms) after the response in the 3-Stimulus Task on average, and 16 ms (±11 ms) after the response in the Radar Task.[7]

We performed a linear mixed effects analysis on the relationship between bump latencies and RTs. To do so, we calculated the elapsed time before the peak of Bump 1, the elapsed time between the peaks of Bumps 1 and 2, and the elapsed time between the peaks of Bumps 2 and 3. We treated these as fixed effects and estimated a random intercept for each participant. We added factors in a stepwise manner beginning from a null model with a random intercept only, and used the likelihood ratio test to compare full and reduced models. If the mental processes reflected by each subsequent bump are necessary for responding, inclusion of the bump in the mixed effects model should account for additional variance in RTs. As shown in Table 6, the likelihood ratio test justified the inclusion of Bump 1 ($\chi_1 = 209.5$) and Bump 2 ($\chi_1 = 201.7$), but not of Bump 3 ($\chi_1 = 0.8$, $p > 0.3$) in the PVT. Conversely, the likelihood ratio test justified the inclusion of Bumps 1, 2, and 3 in the 3-Stimulus

[5] At the level of individual participants, RTs were most strongly correlated with the latencies of Bump 2 (mean $r = 0.39$), followed by Bumps 3 ($r = 0.33$) and 1 ($r = 0.27$).

[6] At the level of individual participants, RTs were also most strongly correlated with the latencies of Bump 3 in the 3-Stimulus Task and the Radar Task ($r = 0.61$ and $r = 0.43$), followed by Bumps 2 ($r = 0.43$ and $r = 0.22$) and 1 ($r = 0.41$ and $r = 0.21$).

[7] To test whether the weaker relationship between the latency of Bump 3 and RTs in the PVT was an artifact of lower signal-to-noise ratio in that task, we computed the amplitude of the third bump at site Pz during each trial. We then divided the mean amplitude of the bump by its standard deviation. The resulting signal-to-noise ratios were higher in the PVT (1.2) and the 3-Stimulus Task (1.2) than in the Radar Task (0.8). This indicates that the lower correlation between the latency of Bump 3 and RTs in the PVT was not caused by a lower signal-to-noise ratio.

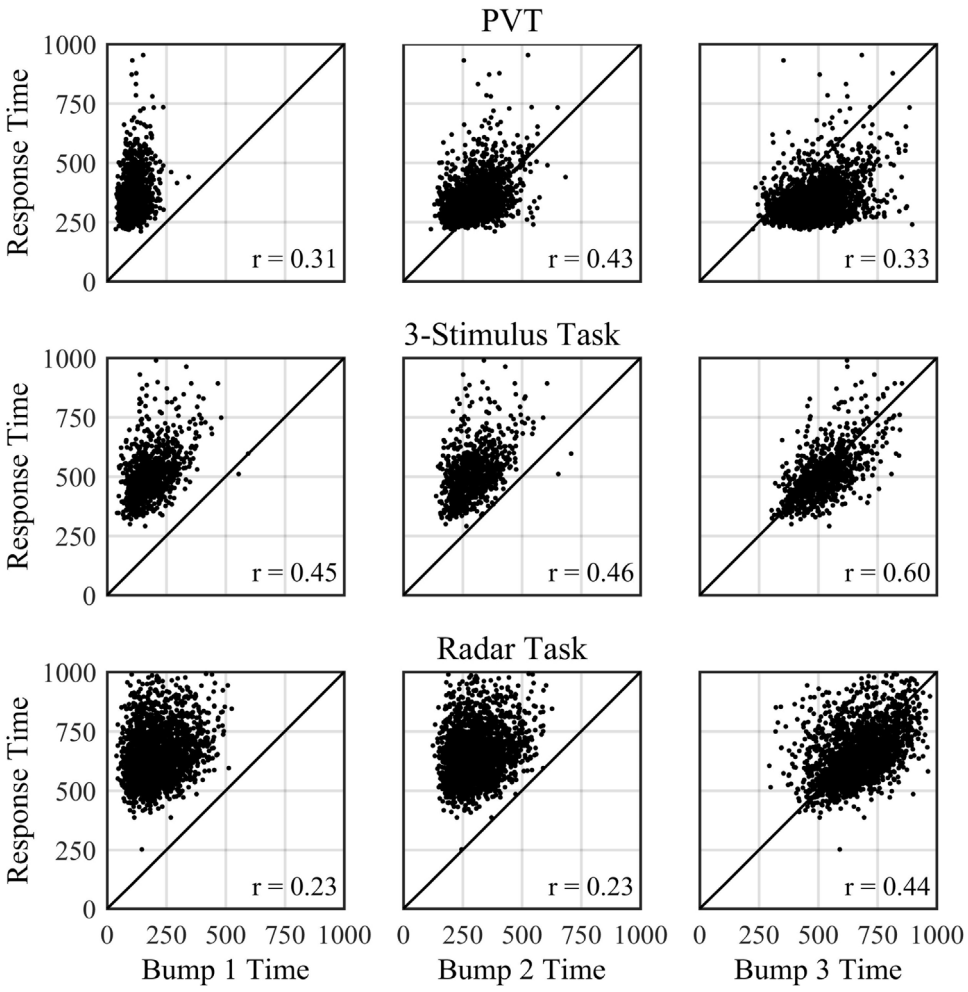

Fig. 7. Scatter plots showing relationship between bump latencies and response times.

Task ($\chi_1 = 177.4$, $\chi_1 = 18.1$, $\chi_1 = 222.3$), as well as in the Radar Task ($\chi_1 = 96.0$, $\chi_1 = 10.0$, $\chi_1 = 303.5$).

P3b latency has been shown to relate to RTs in choice reaction-time tasks (Kutas et al., 1977; Ritter, Simson, & Vaughan, 1983). To determine whether this was the case in our three tasks, we identified the latency of the P3b on a trial-by-trial basis using peak latency and a template matching procedure as described in the methods section. In the PVT, a simple detection task, there was a weak relationship between RTs and the P3b, measured using peak latency ($r = 0.09$) or template matching ($r = 0.18$). In the 3-Stimulus and Radar Tasks, which are choice reaction-time tasks, there was a stronger relationship between RTs and the P3b, measured using peak latency (3-Stimulus: $r = 0.30$; Radar: $r = 0.05$) and template matching (3-Stimulus: $r = 0.29$; Radar: $r = 0.23$).

We performed two additional linear mixed effects analyses to examine the relationship between P3b latency, measured using peak latency or template matching, and RTs. In both analyses, we treated P3b latency as a fixed effect and estimated a random intercept for each participant (Table 6). In the PVT, including P3b latency produced an improvement beyond the null model (peak picking: $\chi_1 = 33.3$, $p < 0.0001$; template matching: $\chi_1 = 60.4$, $p < 0.0001$). In the 3-Stimulus Task, including P3b latency also produced an improvement beyond the null model (peak picking: $\chi_1 = 92.9$, $p < 0.0001$; template matching: $\chi_1 = 74.0$, $p < 0.0001$). Finally, in the Radar Task, including P3b latency measured with template matching, but not peak picking, produced an improvement beyond the null model (peak picking: $\chi_1 = 0.7$, *ns*; template matching: $\chi_1 = 120.6$, $p < 0.0001$).[8]

6. General discussion

The P3b is among the most widely studied components in ERP research. In this paper, we used HSMM-MVPA to analyze data from three variants of the oddball paradigm known to evoke the P3b. We asked whether HSMM-MVPA would reveal a stage in each task corresponding to the P3b, and whether the stage's latency would relate to response times. The onset of the millisecond counter in the PVT produced a bump with a topographical distribution and time course resembling the P3b, as did targets in the 3-Stimulus Task and critical signals in the Radar Task. The latencies of the P3b differed across trial and correlated with RTs in the 3-Stimulus and Radar Tasks, but not in the PVT. We discuss the implications of these findings with respect to the functional significance of the P3b and the use of HSMM-MVPA for mental chronometry in turn.

[8] To allow comparison with previously reported results, we also divided each participants' trials into upper and lower quantiles based on median RTs. We then compared P3b latencies, quantified using the template matching procedure, between the quantiles. P3b latency differed slightly between quantiles in the PVT (355 versus 392 ms, $t(28) = 2.43$, $p < 0.05$), the 3-Stimulus Task (351 versus 404 ms, $t(27) = 3.94$, $p < 0.001$), and the Radar Task (458 versus 513 ms, $t(27) = 5.74$, $p < 0.001$).

6.1. Functional significance of the P3b

A long standing theory is that the P3b reflects context updating (Donchin, 1981; Polich, 2007). Incoming stimuli are attended to and compared against a stimulus representation in memory. If no change is detected, the stimulus context is maintained and only sensory components are evoked. If change is detected, however, the stimulus context is updated and a P3b is evoked, reflecting the corresponding neural activity following the earlier sensory components. This account explains why manipulations that exacerbate sensory processing and prolong stimulus categorization delay P3b latencies. The question of how the P3b relates to responding remains controversial, however. According to the strategic view, the P3b reflects an overarching process not directly relevant to immediate responses (Donchin & Coles, 1988). According to the tactical view, the P3b bridges stimulus categorization and response preparation (Verleger et al., *in press*; Verleger et al., 2005). Both accounts ascribe the P3b a role in stimulus processing. They differ, however, in whether the P3b relates strategically to future behavior or tactically to trial-level responses.

In the 3-Stimulus and Radar Tasks, which both required categorizing the stimulus before responding, the P3b bump peaked with responses and correlated with RTs. Critically, the latency of the P3b bump varied across repeated presentations of the same stimuli and was predictive of RTs. This strongly suggests that the process reflected by the P3b bump contributes tactically to trial-level responses in the 3-Stimulus Task and the Radar Task. In the PVT, which required simple detection, the P3b came about 150 ms after responses and did not correlate with RTs after controlling for the latencies of earlier bumps. This indicates that the process reflected by the P3b bump did not contribute tactically to trial-level responses in the PVT. The central finding pertaining to the P3b bump, and the P3b itself, is that it plays a tactical role in tasks that require stimulus categorization, but not in tasks that require simple detection.

Ouyang et al. (2011, 2013) recently proposed a novel technique called residue iteration decomposition (RIDE) to separate the EEG into stimulus-locked, response-locked, and central component clusters. The technique involves iteratively estimating the latencies of central clusters during single trials, and iteratively improving estimates of the stimulus-locked, central, and response-locked components. Central components bridge stimulus- and response-related processes. Verleger et al. (*in press*) applied RIDE to EEG data from a task with conditions that involved rare and frequent go and no-go responses. Participants responded more slowly to rare-go stimuli than to frequent-go stimuli. RIDE revealed a central cluster focused over the middle of the scalp in the rare-go condition but not in the frequent-go condition. Verleger et al. interpreted these results in terms of a two-process account; controlled processing in the case of rare-go stimuli produced a P3b that bridged stimulus categorization and responding, whereas automatic processing in the case of frequent-go stimuli allowed participants to bypass categorization and respond to stimulus onset. Our results are entirely consistent with this interpretation. Because controlled processing was needed in the 3-Stimulus Task and the Radar Task, the latency of the P3b predicted RTs in those tasks. Alternatively, because automatic processing was possible in the PVT, the P3b followed responses and did not relate to RTs in that task.

Twomey et al. (2015) recently proposed that the P3b reflects accumulation of evidence toward an action-triggering threshold. They found that although P3b amplitude varied with decision difficulty and response speed in stimulus-locked waveforms, P3b amplitude was equivalent across conditions and for fast and slow responses in response-locked waveforms. Twomey et al. (2015) interpreted this as meaning that the P3b builds to a fixed decision criterion after which a response occurs. Our results are partially consistent with their account. The latency but not the amplitude of the P3b bump varied with RTs in the 3-Stimulus and Radar Tasks, suggesting that the P3b builds to roughly the same value in all target trials ending in a response.[9] If the P3b purely reflected accumulation of evidence toward an action criterion, however, we would also expect for it to precede and be predictive of responses in the PVT, which it was not. These results could potentially be reconciled if the P3b built to a decision about the identity of a stimulus, rather than the response. Such a decision is required to respond in choice tasks (e.g., the 3-Stimulus and Radar Tasks) but not simple RT tasks (e.g., the PVT).

6.2. Mental chronometry

Multiple studies have demonstrated a relationship between P3b latency and RTs. Some of these involve blocked designs where mean P3b latency and RTs are computed in multiple conditions and compared to one another (McCarthy & Donchin, 1981; Ritter et al., 1982; for a review, see Verleger, 1997). The general finding is that P3b latency is more sensitive to manipulations that affect stimulus versus response processing. Other studies divide participants' RTs into quantiles, and compare the amplitude and latency of the P3b averaged across trials that make up quantiles (Friedman, 1984; Holm, Ranta-aho, Sallinen, Karjalainen, & Müller, 2006; Kida et al., 2003; Morgan, Luu, & Tucker, 2016; Ramchurn et al., 2014; Roth et al., 1978). These studies draw on the fact that P3b latencies to repeated presentations of the same stimuli vary by tens of milliseconds within individuals (Nishida et al., 1997). The typical finding from these studies is that the P3b amplitude is diminished and its latency is longer in trials with slower responses. Far fewer studies have compared variability in P3b latency and RTs on a trial-by-trial basis. In one of the few studies to do so, Kutas et al. (1977) used template matching to quantify P3b latency in each trial (see also Pfefferbaum et al., 1980a,b; Ritter, Simson, & Vaughan, 1972). Regardless of whether instructions stressed speed or accuracy, P3b latencies correlated with RTs when participants responded correctly.[10]

The correlations we calculated were based on single trial responses. In the 3-Stimulus and Radar Tasks, the latencies of the P3b bump correlated with RTs ($r = 0.60$ and $r = 0.44$, respectively). These correlations were statistically significant for most participants in both tasks (25 of 28 in the 3-Stimulus Task and 21 of 28 in the Radar Task). P3b latencies measured using peak amplitude and template matching, two classic procedures (Woody, 1967), were only weakly to moderately predictive of RTs during individual trials. Consistent with earlier studies, P3b latencies were predictive of RTs at coarser levels of aggregation, however. For example, the correlation between group-averaged P3b peak latencies and RTs across the three tasks was high ($r = 0.90$), and the P3b peaked earlier during the fastest 50% of trials in all tasks. However, as compared to P3b latencies, HSMM-MVPA provided a far more reliable indicator of the timing of mental events culminating in responses during individual trials. The performance of HSMM-MVPA is even more impressive given that the data were extensively filtered (i.e., a 3.5 Hz low-pass filter) before applying peak picking and template matching, yet HSMM-MVPA still accounted for far more variance in RTs.

[9] We binned trials based on median RTs, and compared voltages at the peak latencies of the P3b bumps for each quantile. Mean voltages over the parietal region at the peak of the third bump were similar for the fast and slow response quantiles in the PVT (16.6 μV versus 17.2 μV), the 3-Stimulus task (19.7 μV versus 18.8 μV), and the Radar task (13.5 μV versus 12.5 μV).

[10] Quantifying the relationship between response times and P3b latencies in error trials is complicated by the fact that error responses elicit a late error positivity that overlaps the P3b (Verleger et al., 2005).

HSMM-MVPA is more reliable than peak picking and template matching for two reasons. First, HSMM-MVPA detects bumps with multi-dimensional distributions of activity across the entire scalp, whereas peak picking and template matching detect a feature (i.e., maximum voltage or maximum cross-covariance with a template) over a single sensor or region. The presence of more features in the case of HSMM-MVPA may increase its robustness against noise (cf. Jaskowski & Verleger, 2000; Tomberg, 1995). Second, HSMM-MVPA uses information from all trials to estimate the gamma distributions that influence the durations of the flats preceding the bumps. Single trial estimates of bump timing, then, take advantage of constraints from all other trials. Peak picking and template matching are typically applied over a fixed window. Latency estimates in a given trial are not influenced by estimates from other trials. The added constraint to minimize differences in bump latencies between trials in HSMM-MVPA further increases its robustness against noise.

In addition to being more robust against noise, HSMM-MVPA has another significant advantage beyond peak picking and template matching: it can be used to discover the number of bumps in the EEG, and to temporally localize them on a trial-by-trial basis. HSMM-MVPA detected significant mental events besides the transition from stimulus- to response-related processing signaled by the appearance of the P3b. Across all conditions of the experiments, HSMM-MVPA also detected bumps related to the N1 and the N2b. Their latencies, though more weakly correlated with RTs (cf. Ouyang, Schacht, Zhou, & Sommer, 2013; Verleger et al., 2016), nonetheless accounted for trial-by-trial variability in RTs (Fig. 7). This replicates other reports that the latencies of early components are predictive of RTs (Kida et al., 2003; Ribeiro, Paiva, & Castelo-Branco, 2016).

According to Sternberg's additive-factors logic, RTs reflect the cumulative duration of successive processing stages beginning from stimulus onset (1969). The latencies of later stages necessary for responding are more strongly associated with RTs because they encompass the cumulative variability of all earlier stages. Successively later bumps in the three experiments tended to be more strongly associated with RTs. The one exception was the PVT, where the second bump was most strongly associated with responses. This indicates that the third (P3b) bump reflects a mental process that is not needed to respond in the PVT. The ability to isolate multiple stage durations from EEG enables more refined measurement of how experimental factors impact specific mental processes than cumulative measures like RT allow (Anderson et al., 2016; Zhang et al., *in press*).

The trial-level correlations between bump latencies and RTs, though high, were not perfect. Remaining variance in RTs could relate to limitations in HSMM-MVPA's ability to identify precisely when neural events occurred during a trial. In earlier simulation studies where HSMM-MVPA was used to recover known locations of bumps from synthetic data on a trial-by-trial basis, root-mean-squared deviations ranged from 50 to 100 ms. Based on the variance in RTs from the 3-Stimulus and Radar Tasks, an absolute estimation error on the order of 100 ms would account for nearly all of the unexplained variance in the trial-level estimates of P3b bump latencies and RTs. Alternatively, RTs could be influenced by additional processes that occurred subsequent to the P3b bump such as checking and motor planning. Notwithstanding these limitations, HSMM-MVPA provides a sensitive measure of the timing of neural events at the level of single trials.

6.3. Early sensory components

Our main focus was on the P3b, yet HSMM-MVPA also detected bumps related to the N1 and the N2b. Unlike the P3b, these bumps appeared in all conditions of the tasks. The topographical distribution and latency of the N2b bump was relatively consistent across experiments and conditions. This component is thought to be involved in stimulus classification, but is more sensitive to perceptual features of the stimulus than the P3b is (Folstein & Van Petten, 2008). For example, Daffner et al. (2000) presented participants with visual stimuli that were unfamiliar or familiar. The N2b amplitude depended on whether the stimulus was unfamiliar or familiar, whereas the P3b amplitude depended only on whether the stimulus was a target or a distractor. It is sensible that responses in the PVT occurred during perceptual processing and before stimulus categorization, as detection is all that was needed to respond in that task.

The amplitude of the N2b is sometimes larger for infrequent non-target stimuli than for target stimuli (Folstein & Van Petten, 2008), however the N2b did not vary with stimulus type or frequency in our experiments. In other studies that have used the 3-Stimulus Task, the N2b is greater for infrequent stimuli that are highly perceptually deviant (Demiralp et al., 2001; Polich & Comerchero, 2003). However, when infrequent stimuli are not perceptually deviant, as in our experiments, differences in N2b amplitude between conditions are small or absent (Comerchero & Polich, 1999; Polich & Comerchero, 2003; Sawaki & Katayama, 2008).

A bump corresponding to the N1 clearly appeared between 160 and 190 ms in the 3-Stimulus and Radar Tasks, but not the PVT. The N1 is evoked by the presentation of a stimulus at an attended location, and is greater in choice than in simple RT tasks (Vogel & Luck, 2000). The presence of an N1 bump between 160 and 190 ms in the 3-Stimulus and Radar Tasks is consistent with the fact that they involve choice responses whereas the PVT involves simple detection only.

A bump with a topography very similar to the N1 appeared around 80 ms in the PVT (Fig. 5). When we binned trials according to the latency of the first bump and plotted waveforms over the parietal region, an 8 to 10 Hz oscillation originating during the prestimulus interval and extending to about 150 ms after stimulus onset became evident. The first bump in the 3-Stimulus Task and the first bump in the Radar Tasks were extremely similar to the first bump in the PVT, but they occurred about 100 ms later. This raises the possibility that the first bump in the PVT, and the first bumps in the 3-Stimulus and Radar Tasks reflect consecutive cycles of an alpha oscillation. Consistent with this idea, when we constrained the HSMM-MVPA to estimate a bump at 80 ms in the 3-Stimulus and Radar Tasks, the constrained bump exhibited a posterior negativity and was nearly identical to the first bump from the PVT. Binning trials according to the latency of the constrained bump again revealed an alpha oscillation extending from the prestimulus interval until about 180 ms after stimulus onset.

The question of whether ERPs arise from the superposition of event-locked signals atop ongoing EEG activity, temporal resetting of the phase structure of ongoing EEG activity, or some combination of both is basic to EEG research (Basar, 1980; Van der Lubbe, Szumska, & Fajkowska, *in press*; Vaughan, 1969; Yeung et al., 2004, 2007). Our results indicate that activity driving the first bump in the PVT at 80 ms (and the temporally constrained bumps in the 3-Stimulus and Radar Tasks at 80 ms) originates during the prestimulus interval (Fig. 6). Stimulus presentation did not cause the phase of oscillation to become aligned by 80 ms across trials – otherwise, a negative-going potential would be evident around 80 ms in the ERPs. The oscillation's amplitude did appear to increase following stimulus presentation, however. As such, these results suggest that stimulus-presentation increased the amplitude of an ongoing alpha oscillation, but did not induce phase realignment by 80 ms.

These unanticipated outcomes pertaining to the N1 are consistent with a study by Makeig et al. (2002) where alpha ringing was observed over the parietal scalp during a visual attention task.

Makeig et al. (2002) reported peak N1 latency around 200 ms. By plotting the subset of trials with greatest power in the alpha frequency range, they showed additional negative troughs occurring at 100 ms intervals preceding and following the N1. This suggests that the N1 may arise from increased power and/or phase synchronization in the alpha frequency band following stimulus presentation (see also Klimesch et al., 2004). Interestingly, although HSMM-MVPA models neural signals as half-cycle sine waves, it revealed an alpha oscillation by detecting the increased amplitude of the first negative cycle following stimulus onset. Because the locations of bumps are allowed to vary on a trial-by-trial basis, HSMM-MVPA detected the first negative cycle despite the fact that it was not stimulus locked.

The power of EEG at each frequency is comprised of evoked and induced power. Evoked power is an event-related change in the power of an oscillation whose period is phase-synchronized to the event, whereas induced power is event-related but not phase-synchronized. The oscillation we detected was concentrated in the lower alpha band (8–10 Hz). The oscillation was not stimulus locked, and would thus appear as a combination of sustained power from the baseline period and additional induced power. Van der Lubbe et al. (*in press*) recently performed a wavelet analysis of visual potentials to faces. They found that the amplitude of the N1 related to an increase in evoked power in the theta band. The apparent discrepancy between those results and ours could relate to either of two factors. First, the oscillation we observed was not stimulus locked and would thus appear as a change in induced rather than evoked power. Second, our analysis does not rule out the possibility of concurrent oscillations in other frequency bands. In the absence of amplitude differences, HSMM-MVPA will tend to detect alpha oscillations because (1) the temporal duration of bumps (50 ms) coincides with a half-cycle of an alpha oscillation, and (2) the greater durations between adjacent peaks of slower oscillations will cause them to appear at less predictable times on a trial-by-trial basis.

6.4. Conclusion

The P3b has been extensively studied, and has been used to support inferences about the timing and types of mental processes that occur in a range of tasks. We applied a novel methodology, HSMM-MVPA, to data from three paradigms commonly used to study the P3b. The methodology revealed distinct neural signals corresponding to the P3b and earlier visual components. The latencies of these bumps, and particularly of the P3b bump, correlated with single trial RTs when stimulus categorization was needed to respond. These results provide new information about the functional significance of the P3b, and demonstrate the utility of a new approach to EEG-based mental chronometry.

Acknowledgements

This research was performed while MMW held a National Research Council Research Associateship Award with the Air Force Research Laboratory's (AFRL) Cognitive Models and Agents Branch. MMW was also supported by ONR grant N00014-15-1-2151. The authors thank AFRL for supporting this research. Special thanks to Ashley Chaffin and Rachel Boyd for their technical support in conducting this research, and Qiong Zhang for her comments on an earlier version of this manuscript. The views expressed in this paper do not represent the official opinion of the United States government or the Air Force.

References

Anderson, J. R., Zhang, Q., Borst, J. P., & Walsh, M. M. (2016). The measurement of processing stages: Extension of Sternberg's method. *Psychological Review*.
Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* USA: Oxford University Press.
Basar, E. (1980). *EEG brain dynamics: Relation between EEG and brain evoked potentials*. Amsterdam: Elsevier.
Basner, M., & Dinges, D. F. (2011). Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep*, *34*, 581–591.
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.
Coles, M. G., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(5), 529–553.
Comerchero, M. D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology*, *110*(1), 24–30.
Conroy, M. A., & Polich, J. (2007). Normative variation of P3a and P3b from a large sample: Gender, topography, and response time. *Journal of Psychophysiology*, *21*(1), 22–32.
Croft, R. J., Gonsalvez, C. J., Gabriel, C., & Barry, R. J. (2003). Target-to-target interval versus probability effects on P300 in one-and two-tone tasks. *Psychophysiology*, *40*(3), 322–328.
Daffner, K. R., Mesulam, M. M., Scinto, L. F., Calvo, V., Faust, R., & Holcomb, P. J. (2000). An electrophysiological index of stimulus unfamiliarity. *Psychophysiology*, *37*(06), 737–747.
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.
Demiralp, T., Ademoglu, A., Comerchero, M., & Polich, J. (2001). Wavelet analysis of P3a and P3b. *Brain Topography*, *13*(4), 251–267.
Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–38.
Dien, J. (2010). Evaluating two-step PCA of ERP data with geomin, infomax, oblimin, promax, and varimax rotations. *Psychophysiology*, *47*, 170–183.
Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers*, *17*(6), 652–655.
Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(3), 355–425.
Donchin, E. (1981). Surprise!. . . surprise? *Psychophysiology*, *18*(5), 493–513.
Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, *14*(5), 456–467.
Duncan-Johnson, C. C. (1981). Young psychophysiologist award address: 1980. P300 latency: A new metric of information processing. *Psychophysiology*, *18*, 207–215.
Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*(1), 152–170.
Friedman, D. (1984). P300 and slow wave: The effects of reaction time quartile. *Biological Psychology*, *18*(1), 49–71.
Gonsalvez, C. J., Barry, R. J., Rushby, J. A., & Polich, J. (2007). Target-to-target interval, intensity, and P300 from an auditory single-stimulus task. *Psychophysiology*, *44*, 245–250.
Hitchcock, E. M., Dember, W. N., Warm, J. S., Moroney, B. W., & See, J. E. (1999). Effects of cueing and knowledge of results on workload and boredom in sustained attention. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *41*(3), 365–372.
Holm, A., Ranta-aho, P. O., Sallinen, M., Karjalainen, P. A., & Müller, K. (2006). Relationship of P300 single-trial responses with reaction time and preceding stimulus sequence. *International Journal of Psychophysiology*, *61*(2), 244–252.
Hopf, J. M., Vogel, E., Woodman, G., Heinze, H. J., & Luck, S. J. (2002). Localizing visual discrimination processes in time and space. *Journal of Neurophysiology*, *88*(4), 2088–2095.
Jaśkowski, P., & Verleger, R. (2000). An evaluation of methods for single-trial estimation of P3 latency. *Psychophysiology*, *37*(2), 153–162.
Kamzanova, A. T., Kustubayeva, A. M., & Matthews, G. (2014). Use of EEG workload indices for diagnostic monitoring of vigilance decrement. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *56*(6), 1136–1149.
Katayama, J. I., & Polich, J. (1996). P300 from one-, two-, and three-stimulus auditory paradigms. *International Journal of Psychophysiology*, *23*(1), 33–40.
Kida, T., Nishihira, Y., Hatta, A., Wasaka, T., Nakata, H., Sakamoto, M., & Nakajima, T. (2003). Changes in the somatosensory N250 and P300 by the variation of reaction time. *European Journal of Applied Physiology*, *89*(3–4), 326–330.
Klimesch, W., Schack, B., Schabus, M., Doppelmayr, M., Gruber, W., & Sauseng, P. (2004). Phase-locked alpha and theta oscillations generate the P1–N1 complex and are related to memory performance. *Cognitive Brain Research*, *19*, 302–316.
Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, *197*(4305), 792–795.
Lim, J., & Dinges, D. F. (2008). Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences*, *1129*, 305–322.
Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Magliero, A., Bashore, T. R., Coles, M. G., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology*, *21*(2), 171–186.
Makeig, S., Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2002). Dynamic brain sources of visual evoked responses. *Science*, *295*(5555), 690–694.
McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science*, *211*(4477), 77–80.
Mertens, R., & Polich, J. (1997). P300 from a single-stimulus paradigm: Passive versus active tasks and stimulus modality. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *104*(6), 488–497.
Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*, *26*(1–3), 3–67.
Morgan, K. K., Luu, P., & Tucker, D. M. (2016). Changes in P3b latency and amplitude reflect expertise acquisition in a football visuomotor learning task. *PloS one*, *11*, e0154021.
Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus–norepinephrine system. *Psychological Bulletin*, *131*(4), 510–532.
Nishida, S., Nakamura, M., Suwazono, S., Honda, M., & Shibasaki, H. (1997). Estimate of physiological variability of peak latency in single sweep P300. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *104*(5), 431–436.
O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729–1735.
Ouyang, G., Herzmann, G., Zhou, C., & Sommer, W. (2011). Residue Iteration Decomposition (RIDE): A new method to separate ERP components on the basis of latency variability in single trials. *Psychophysiology*, *48*, 1631–1647.
Ouyang, G., Schacht, A., Zhou, C., & Sommer, W. (2013). Overcoming limitations of the ERP method with Residue Iteration Decomposition (RIDE): A demonstration in go/no-go experiments. *Psychophysiology*, *50*(3), 253–265.
Pfefferbaum, A., Ford, J. M., Roth, W. T., & Kopell, B. S. (1980a). Age differences in P3-reaction time associations. *Electroencephalography and Clinical Neurophysiology*, *49*, 257–265.
Pfefferbaum, A., Ford, J. M., Roth, W. T., & Kopell, B. S. (1980b). Age-related changes in auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology*, *49*(3–4), 266–276.
Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, *9*, 456–479.
Polich, J., & Comerchero, M. D. (2003). P3a from visual stimuli: Typicality, task, and topography. *Brain Topography*, *15*(3), 141–152.
Polich, J., & Heine, M. R. (1996). P300 topography and modality effects from a single-stimulus paradigm. *Psychophysiology*, *33*(6), 747–752.
Polich, J., Eischen, S. E., & Collins, G. E. (1994). P300 from a single auditory stimulus. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *92*(3), 253–261.
Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148.
The R Core Team. (2016). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria: The R Core Team. URL https://www.R-project.org/
Ramchurn, A., de Fockert, J. W., Mason, L., Darling, S., & Bunce, D. (2014). Intraindividual reaction time variability affects P300 amplitude rather than latency. *Frontiers in Human Neuroscience*, *8*, 557.
Ribeiro, M. J., Paiva, J. S., & Castelo-Branco, M. (2016). Spontaneous fluctuations in sensory processing predict within-subject reaction time variability. *Frontiers in Human Neuroscience*, 10.
Ritter, W., Simson, R., & Vaughan, H. G. (1972). Association cortex potentials and reaction time in auditory discrimination. *Electroencephalography and Clinical Neurophysiology*, *33*(6), 547–555.
Ritter, W., Simson, R., Vaughan, H. G., & Macht, M. (1982). Manipulation of event-related potential manifestations of information processing stages. *Science*, *218*, 909–911.
Ritter, W., Simson, R., & Vaughan, H. G. (1983). Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks. *Psychophysiology*, *20*(2), 168–179.
Roth, W. T., Ford, J. M., & Kopell, B. S. (1978). Long-latency evoked potentials and reaction time. *Psychophysiology*, *15*(1), 17–23.
Saville, C. W., Lancaster, T. M., Davies, T. J., Toumaian, M., Pappa, E., Fish, S., . . . & Klein, C. (2015). Elevated P3b latency variability in carriers of ZNF804A risk allele for psychosis. *NeuroImage*, *116*, 207–213.
Sawaki, R., & Katayama, J. I. (2008). Distractor P3 is associated with attentional capture by stimulus deviance. *Clinical Neurophysiology*, *119*(6), 1300–1309.
Smulders, F. T., Kok, A., Kenemans, J. L., & Bashore, T. R. (1995). The temporal selectivity of additive factor effects on the reaction process revealed in ERP component latencies. *Acta Psychologica*, *90*(1), 97–109.
Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401.
Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276–315.
Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(3700), 1187–1188.
Tarkka, I. M., & Stokic, D. S. (1998). Source localization of P300 from oddball, single stimulus, and omitted-stimulus paradigms. *Brain Topography*, *11*(2), 141–151.
Tomberg, C. (1995). New application of the Z estimator to identify the cognitive P300 in non-averaged human brain potentials. *Neuroscience Letters*, *184*, 153–156.
Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, *42*(1), 1636–1643.
Van der Lubbe R. H. J., Szumska I., & Fajkowska M. (*in press)* Two sides of the same coin: ERP and wavelet analysis of visual potentials evoked and induced by task-relevant faces. *Advances in Cognitive Psychology*.
Vaughan, H. G. (1969). The relationship of brain activity to scalp recordings of event-related potentials. In E. Donchin, & D. B. Lindsley (Eds.), *Average evoked potentials: Methods, results, and evaluations*. Washington, DC: U.S. Government Printing Office.
Verleger R., Grauhan N., & Śmigasiewicz K. (*in press*) Is P3 a strategic or a tactical component? Relationships of P3 sub-components to response times in oddball tasks with go, no-go and choice responses. *NeuroImage*.
Verleger, R., Jaśkowski, P., & Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *Journal of Psychophysiology*, *19*(3), 165–181.
Verleger, R. (1988). Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of P3. *Behavioral and Brain Sciences*, *11*, 343–356.
Verleger, R. (1997). On the utility of P3 latency as an index of mental chronometry. *Psychophysiology*, *34*, 131–156.
Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*(02), 190–203.
Walhovd, K. B., & Fjell, A. M. (2001). Two-and three-stimuli auditory oddball ERP tasks and neuropsychological measures in aging. *Neuroreport*, *12*(14), 3149–3153.
Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: An electrical sign of sensorimotor association and expectancy in the human brain. *Nature*, *203*, 380–384.
Witkowski, S., Trujillo, L. T., Sherman, S. M., Carter, P., Matthews, M. D., & Schnyer, D. M. (2015). An examination of the association between chronic sleep restriction and electrocortical arousal in college students. *Clinical Neurophysiology*, *126*(3), 549–557.
Woody, C. D. (1967). Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Medical and Biological Engineering*, *5*, 539–553.
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*(1), 39–49.
Yu, S. Z. (2010). Hidden semi-Markov models. *Artificial Intelligence*, *174*(2), 215–243.
Zhang, Q., Walsh, M. M., & Anderson, J. R. (in press) The effects of probe similarity on retrieval and comparison processes in associative recognition. *Journal of Cognitive Neuroscience*.