Psychophysiology, 52 (2015), 460–471. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/psyp.12378

ERP and pupil responses to deviance in an oddball paradigm

SIRI-MARIA KAMP[a] AND EMANUEL DONCHIN[b]

[a]Department of Psychology, Saarland University, Saarbrücken, Germany
[b]Department of Psychology, University of South Florida, Tampa, Florida, USA

Abstract

We investigated the relationship between, and functional significance of, P300, novelty P3, and the pupil dilation response (PDR). Subjects categorized stimuli including (a) words of a frequent category, (b) words of an infrequent category (14%), and (c) pictures of the frequent category ("novels"; 14%). The P300 and novelty P3 were uncorrelated with the PDR and differed in their response to experimental manipulation. Therefore, although the three physiological responses often co-occur, they appear to each manifest a distinct function: The PDR may be more closely linked to aspects of behavioral responding than the event-related potentials. Within participants, P300 and PDR latencies accounted for unique portions of the reaction time variance, and amplitudes of all three responses were larger for stimuli recalled on a subsequent test, compared to not recalled. We discuss the possibility that all three responses reflect norepinephric input from the locus coeruleus.

Descriptors: EEG/ERP, Pupillometry, Cognition

Humans and other animals preferentially attend to and remember novel events. In line with its biological significance, novelty elicits a set of autonomic responses collectively labeled the "orienting reflex" (for reviews, see Kimmel, 1979; Sokolov, 1963). Furthermore, many event-related potential (ERP) components, including the mismatch negativity (MMN), N2, novelty P3, P300, and N400, are sensitive to novelty (Donchin, Spencer, & Dien, 1997; Fabiani, 2006). In particular, the eliciting conditions of the P300 show striking similarities to a temporary dilation of the pupil (the pupil dilation response; PDR; see Donchin et al., 1984; Nieuwenhuis, De Geus, & Aston-Jones, 2011). Their co-occurrence raises the question whether P300 and PDR manifest the same neurocognitive operation invoked upon the encounter with unexpected events (such as response facilitation), or whether their functions are different (such as responding vs. episodic encoding).

To investigate this, we recorded ERPs and pupil size simultaneously in an oddball task expected to elicit both a P300 and a PDR. We additionally took the novelty P3 into consideration, which overlaps with the P300. If the responses index the same cognitive function, they should vary analogously with experimental manipulation and with behavior. In turn, if their cognitive functions are different, this variance should be dissociable. More specifically, we asked whether each response is integrated into the stimulus-response stream, as indicated by a correlation of its latency with reaction time (RT), or indexes episodic encoding, indicated by a correlation of its amplitude with subsequent recall. Our hypotheses were guided by previous findings as well as theoretical accounts, in particular the context updating hypothesis of the P300 and the locus coeruleus norepinephrine (LC-NE) theory of P300 and PDR.

The P300

The P300 (Sutton, Braren, Zubin, & John, 1965) is a positivity that is largest at parietal scalp sites and peaks at least 300 ms after the eliciting event. It is often studied in the oddball paradigm, in which events in a sequence can be classified into a rarely and a frequently occurring category. Events of the rare category elicit a P300. The context updating hypothesis (Donchin, 1981; Donchin & Coles, 1988) assumes that a mental schema of goal-relevant information (Miller, Galanter, & Pribram, 1960) must be "updated" when new information conflicts with expectations derived from it—the detection of the need for context updating is manifested in the P300. The P300 is assumed not to be a stage in a serial process that leads from stimulus to response, but to be part of a parallel processing stream affecting future behaviors.

P300 and RT. P300 latency often correlates with RT (for a review, see Verleger, 1997). This is consistent with the context updating theory, because neither can the schema be updated, nor can an accurate response be made, before the stimulus has been

We would like to thank Margaret M. Bradley, who in the context of a Research Fellowship Training Award from the Society for Psychophysiological Research provided S. Kamp with excellent training on eye tracking and pupillometry. We are also grateful to Glen R. Forester, Anthony R. Murphy, Ty Brumback, as well as the research assistants of the Cognitive Psychophysiology Laboratory at the University of South Florida for help with data collection and other research assistance. Finally, we thank Michael Weigl and our reviewers for helpful comments on this manuscript. S. Kamp was supported by a Signature Research Doctoral Fellowship awarded by the University of South Florida during completion of this research project.

Address correspondence to: Siri-Maria Kamp, Saarland University, International Research Training Group "Adaptive Minds," Department of Psychology, Campus, Building A2.4, D-66123 Saarbrücken, Germany.
E-mail: siri.kamp@uni-saarland.de

PSYCHOPHYSIOLOGY

classified as deviant. Importantly, under some circumstances the two measures are dissociable, for example when response speed is emphasized over accuracy encouraging responding before full stimulus evaluation (Kutas, McCarthy, & Donchin, 1977). Furthermore, response conflict (such as in the Stroop task) induces a RT cost but does not affect P300 latency (Duncan-Johnson & Kopell, 1981; Ilan & Polich, 1999; see also Magliero, Bashore, Coles, & Donchin, 1984; McCarthy & Donchin, 1981), suggesting that P300 does not index responding directly. Nevertheless, a different view of the P300 that is based on its putative neural generators emphasizes a role in behavioral responding to stimuli or events (Nieuwenhuis, Aston-Jones, & Cohen, 2005; Nieuwenhuis et al., 2011): The P300 may reflect synchronized norepinephric (NE) cortical input from the locus coeruleus (LC), which plays a role in sensory and behavioral adaptation (e.g., Berridge & Waterhouse, 2003). In this view, the P300 facilitates responding to goal-relevant stimuli.

P300 and encoding. The context updating hypothesis links the P300 to episodic encoding, thus predicting that larger P300 amplitudes should be associated with a higher probability of later recalling the eliciting event. Indeed, deviant stimuli, which lead to enhanced memory (Von Restorff, 1933), elicit larger P300 amplitudes when they are subsequently recalled, compared to not-recalled items. This P300 "subsequent memory effect" is observed when encoding is incidental (Fabiani, Karis, & Donchin, 1986) or based on simple rehearsal (Fabiani & Donchin, 1995; Fabiani, Karis, & Donchin, 1990; Kamp, Brumback, & Donchin, 2013; Karis, Fabiani, & Donchin, 1984), but not when encoding focuses on interitem associations (Fabiani et al., 1990; Karis et al., 1984). This suggests a relationship of the P300 to encoding that depends on strategic processes.[1]

The Novelty P3

The second positivity of interest is the frontocentrally distributed novelty P3. It exhibits substantial spatiotemporal overlap with the P300, but can be disentangled from it by use of principal component analysis (PCA; Spencer, Dien, & Donchin, 1999, 2001). The novelty P3 is elicited by some stimuli that also elicit a P300, including task-irrelevant presentations of salient stimuli ("novels") in oddball tasks (Courchesne, Hillyard, & Galambos, 1975; Friedman, Simpson, & Hamberger, 1993). However, perceptual deviance is more crucial than task irrelevance to elicit a novelty P3 (Cycowicz & Friedman, 1999, 2004; Gaeta, Friedman, & Hunt, 2003).

Novelty P3 and RT. Due to similar eliciting conditions to the orienting reflex (Sokolov, 1963), Courchesne et al. (1975) suggested that the novelty P3 reflects an "orienting response" (for a review, see Friedman, Cycowicz, & Gaeta, 2001). Alternatively, it may index response inhibition (Goldstein, Spencer, & Donchin, 2002), because the typical paradigm requires a response to all stimuli but the novels, and because a morphologically similar positivity is elicited by no-go stimuli in a go/no-go paradigm (e.g., Pfefferbaum, Ford, Weller, & Kopell, 1985). However, adding a response requirement does not abolish the novelty P3 (Cycowicz & Friedman, 2004; Gaeta et al., 2003), so its role in response adaptation may be a more general one. The relationship between novelty P3 latency and RT on the same trial has yet to be studied; modifying the novelty oddball paradigm such that the novels were classifiable according to the same rule as the other stimuli allowed us to do so.

Novelty P3 and encoding. The role of the novelty P3 in episodic memory is also unclear. On the one hand, Cycowicz and Friedman (1999) reported no correlation between its amplitude and subsequent recognition success. On the other hand, Kamp et al. (2013) found that physically deviant stimuli elicited a novelty P3 subsequent memory effect. Similarly, Butterfield and Mangels (2003) reported that novelty P3 amplitude elicited by negative feedback about the participant's answer to a question predicted subsequent memory for the correct answer.

The Pupil Dilation Response (PDR)

A phasic dilation of the pupil peaking 1–2 s after the stimulus was first reported in the same paradigm in which the P300 was discovered (Friedman, Hakerem, Sutton, & Fleiss, 1973). Like the P300, the PDR is also elicited by infrequent events in oddball paradigms (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Murphy, Robertson, Balsters, & O'Connell, 2011) and is enhanced when an outcome is better or worse than expected (Preuschoff, 't Hart, & Einhäuser, 2011). Furthermore, erroneous responses are followed by a PDR, which, like the P300, is larger for perceived than unperceived errors (Wessel, Danielmeier, & Ullsperger, 2011). Moreover, in situations where N400 amplitude is larger, both the PDR (Kuipers & Thierry, 2011) and the P300 (Arbel, Spencer, & Donchin, 2010) tend to be smaller.

However, in Stroop tasks the PDR (e.g., Laeng, Ørbo, Holmlund, & Miozzo, 2011; Siegle, Steinhauer, & Thase, 2004), but not the P300 (e.g., Duncan-Johnson & Kopell, 1981; Rosenfeld & Skogsberg, 2006), is larger for incongruous word-color combinations. Combined with evidence that the response-locked PDR is sensitive to the complexity of the movement (Richer & Beatty, 1985), this may suggest that the PDR is more closely related to responding than the P300. Supporting a dissociation, in a simple auditory oddball task, Murphy et al. (2011) found no correlation between the amplitudes of PDR and P300.

PDR and RT. Both PDR and P300 may reflect phasic activity in the LC-NE system (Nieuwenhuis et al., 2011). The LC is thought to regulate behavior (Aston-Jones & Cohen, 2005; Aston-Jones, Rajkowski, & Cohen, 1999): When tonic LC activity is low, but phasic activity to task-relevant stimuli is high, the individual is focused on the task and shows good performance. In contrast, high tonic and low phasic activity index low task focus and poor performance. Pupil diameter mirrors this pattern: Small pretrial (baseline) diameters and large phasic PDRs correspond to good performance and vice versa (Gilzenrat et al., 2010; Jepma et al., 2011). Supporting the idea that the PDR indexes processes facilitative of responding, PDR latency is correlated with RT between participants (Nuthmann & Van Der Meer, 2005). However, to our knowledge, no studies have tested this in single trials.

PDR and encoding. The role of the cognitive processes indexed by the PDR in episodic memory is controversial (Goldinger & Papesh, 2012). Thus, Papesh, Goldinger, and Hout (2012) found

1. In order to account for data pointing toward a role of the LC in memory, Nieuwenhuis (2011) revised the LC-NE theory, attributing both immediate action and learning to the P300. The newer version of the theory is thus more compatible with the context updating hypothesis, but it is broader in scope and therefore still not identical.

that the PDR elicited by studied words correlated positively with subsequent recognition confidence. Other studies have reported negative correlations between relative PDR amplitude and subsequent recognition memory for pictures (Kafkas & Montaldi, 2011; Naber, Frässle, Rutishauser, & Einhäuser, 2013). Finally, Võ et al. (2008) found no evidence for a PDR subsequent memory effect.

The Present Study

We investigated, within participants, the relationship among P300, novelty P3, and PDR in a modified novelty oddball task expected to elicit all three physiological responses. If the physiological variables index the same cognitive function (as proposed by the LC-NE theory), then their variance with experimental manipulation should be the same. However, if the measures index different functions, then their experimental variance should be dissociable.

The task required a behavioral response to all stimuli and was followed by memory tests, allowing us to investigate covariance of each physiological response with behavior. Following the LC-NE theory of the PDR, we hypothesized that the PDR indexes a process directly integrated into the stimulus-response stream, its latency thus correlating with RT on the same trial. Following the context updating hypothesis of the P300, we hypothesized that P300 is linked more closely with episodic encoding, predicting that its amplitude will correlate with subsequent recall. In this view, P300 latency should only correlate with RT to the extent that it depends on stimulus evaluation. Therefore, when variance due to the PDR had been statistically accounted for in a regression analysis, the association between P300 latency and RT should break down. Due to a scarcity of relevant prior findings, we did not make strong predictions regarding the covariance of novelty P3 with behavior.

Method

All procedures were in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the University of South Florida. All participants gave informed consent.

Participants

In exchange for partial course credit, 29 students participated. Data from nine participants were excluded because memory performance deviated from the sample by 2 *SD* ($n = 1$), due to data acquisition errors (e.g., failure of eye tracker calibration, $n = 4$), or because excessive artifacts led to fewer than four usable trials for the picture category ($n = 4$). The final sample included 20 participants, aged 18–49 years ($M = 23.5$; $SD = 8$). Six participants were male, three were left-handed, and all were native speakers of English with normal or corrected-to-normal vision.

Stimuli

The stimuli were classifiable according to one of three rules: (1) edible versus inedible (e.g., pizza vs. table), (2) living versus nonliving (e.g., lion vs. pencil), or (3) smaller versus larger than a shoebox (e.g., ant vs. ship; the tasks are henceforth referred to as the "edible," "living," and "size" tasks).

To select pictorial stimuli, 20 participants[2] were asked to provide a label for 40 black-and-white clip art drawings of each category (edible, inedible, etc.), and to rate the naming difficulty on a scale of 1 (*very easy*) to 7 (*very difficult*). We selected the 20 pictures of each category that exhibited the largest between-subject agreement (same label provided by at least 18 participants) and that were rated as easy to name (average rating < 1.8; $M = 1.17$; $SD = 0.12$). For the labels (85%) that were available in the Francis and Kucera (1982) database, word frequency ranged between 1 and 352 occurrences per million ($M = 31.1$; $SD = 72.2$).

Table 1. *Means (SD) for Word Frequency and Length of Words in Each Category*

Task	Category	Word frequency	Word length
Edible			
	Edible	11.5 (14.7)	5.5 (1.4)
	Inedible	12.4 (11.7)	5.5 (1.3)
Living			
	Living	22.5 (25.1)	5.6 (1.4)
	Nonliving	20.5 (21.6)	5.6 (1.4)
Size			
	Smaller than a shoebox	16.4 (19.9)	5.6 (1.6)
	Larger than a shoebox	17 (20)	5.7 (1.5)

Note. Word frequency values represent lemma occurrences per million taken from Francis and Kucera (1982), and word length is given in the number of letters.

For each of the six categories, 136 nouns that were 3 to 9 letters long were drawn from Francis and Kucera (1982; the labels of the selected pictures were excluded). Care was taken to match word frequency and length between the categories of each task (i.e., between the edible and the inedible category, etc.; see Table 1).

Words and pictures were presented in black color on a gray background (RGB: 125, 125, 125). The words were displayed in Arial font spanning 2.8 to 8.6 degrees of visual angle. The largest dimension (width or height) of the pictures was 95 mm, and therefore they spanned 8.6 degrees of visual angle. Pictures exhibited lower luminance than words ($t = 14.46$).

Task and Procedure

Every session took place at 9 am, and the duration did not exceed 2.5 hr. Up to 30 min were dedicated to preparing the electroencephalogram (EEG) recording and calibrating the eye tracker. The experiment was presented with Eprime 2.0 software and completed in a dimly lit room with constant luminance both within and across participant runs. The experiment contained a practice (which was shorter but followed the same structure) and six experimental blocks, each consisting of encoding, free recall, a distraction task, and recognition (Figure 1).

Encoding. Participants completed an oddball task that involved one of three semantic judgments (edible, living, or size), each of which was randomly assigned to two successive blocks. Participants pressed one of two buttons with their left or right hand (response hand was randomized) to classify each stimulus. Participants were informed that memory tests would follow, but were instructed to focus on the encoding task and to respond as quickly and accurately as possible. Since the stimuli were presented in a quick sequence and each stimulus required a response, we assumed that participants would not use interitem associative encoding strategies (this was confirmed by subject reports during debriefing).

2. Participants were 16- to 60-year-old native speakers of English. Eleven were female and nine were male. None participated in the main experiment.

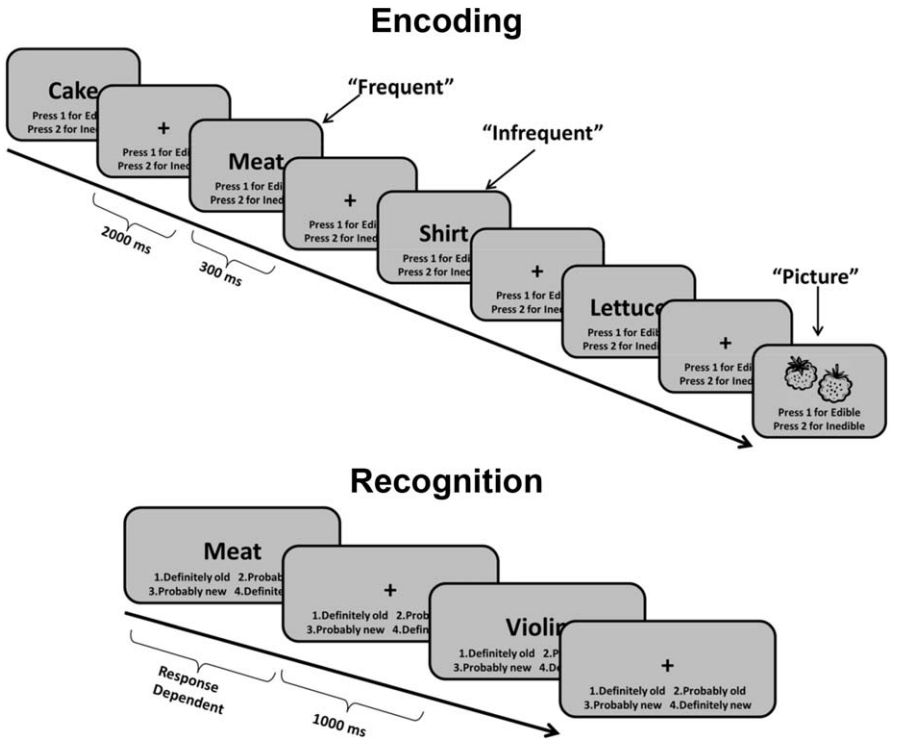

Figure 1. Illustration of the encoding and recognition phases of one experimental block. Shown is a block that included pictures in the encoding sequence (Blocks 2, 4, and 6), and that used the "edible" task (categorize each stimulus into edible vs. inedible). See text for details.

Each of the 73 stimuli in the encoding phase was presented for 300 ms and followed by a fixation cross for 2,000 ms. In Blocks 1, 3, and 5, 63 stimuli (86%) were of the frequent (e.g., edible), and 10 (14%) were of the infrequent (e.g., inedible) category. Blocks 2, 4, and 6 contained 53 frequents (72%), 10 infrequents (14%), and 10 pictures of the frequent category (14%). Which category was infrequent was randomly determined. The pictures served the role of the perceptually deviant novels expected to elicit a novelty P3, but had to be classified according to the same rule as the words and therefore always required the frequent response (however, this was not explicitly mentioned). The first and last three stimuli were always frequents and were not analyzed. Each stimulus was drawn pseudorandomly with the restriction that no two infrequents and no two pictures were presented successively.

Recall. Immediately after each encoding phase, participants were asked to write down, in any order, every word or picture label they remembered. Three minutes later, an experimenter entered the room to take away the recall sheet.

Distraction phase. Pilot data indicated that recognition performance was at ceiling when recognition immediately followed recall. Therefore, participants next completed a 3.5-min long simple oddball task using visually presented letters X and O as stimuli.

Recognition. All infrequents ($n = 10$), all pictures ($n = 10$; only in Blocks 2, 4, and 6), and a random sample of 10 frequents from the encoding phase were presented in random order, along with an equal number of unstudied stimuli drawn from the same pool. Therefore, all stimulus types were equally frequent during the recognition test (however, we will still use the terms frequent and infrequent according to the presentation mode of each category at encoding). This resulted in 60 test trials in the blocks including and 40 test trials in the blocks not including pictures. Participants judged each stimulus on a scale of 1 (*definitely old*) to 2 (*probably old*) to 3 (*probably new*) to 4 (*definitely new*). The stimulus remained on the screen until the response was given, and between two trials a fixation cross was presented for 1,000 ms.

Performance feedback. Upon conclusion of each block, participants were given feedback on their performance during the encoding and recognition phases. The purpose of this feedback was to keep motivation high during the entire experiment. After a break, the next block began.

Measures of Behavioral Performance

We calculated the proportion of correct responses to frequents, infrequents, and pictures during encoding, as well the proportion of stimuli produced during recall. For recognition performance, we analyzed d' as a measure of sensitivity (the ease of distinguishing between an old and a new item), and c to quantify bias (the willingness to respond old vs. new; e.g., Grider & Malmberg, 2008), collapsing across confidence ratings (i.e., combining *definitely* with *probably* responses).[3]

RTs at encoding and test exhibited skewed distributions, so we report subject medians (although analysis of means led to

3. Some participants had a hit rate of 1 or a false alarm rate of 0 for one trial type. Because for such scenarios d' and c are undefined, we added one false alarm, and subtracted one hit for each subject and trial type before calculating d' and c. As the number of pictures was only half the number of frequents and infrequents, for the pictures we added/subtracted 0.5.

equivalent results). Furthermore, we log-transformed RTs for the correlation and regression analyses in order to approximate a normal distribution.

Recording and Preprocessing of Physiological Data

EEG/ERP. The EEG was recorded with a 128 Ag/AgCl electrode EGI system (Eugene, OR), with Cz as the online reference, amplified at 0.1–100 Hz and digitized at a sampling rate of 250 Hz. Using NetStation software, we applied a 20 Hz low-pass filter and mathematically interpolated bad channels. Then, we extracted EEG segments from the encoding phase, including 300 ms before to 1,200 ms after the onset of frequents, infrequents, and pictures. Trials with incorrect responses were excluded. Eye blinks were removed by independent component analysis (ICA) using the ERP PCA toolkit (Dien, 2010). The epochs were rereferenced to the average of the left and right mastoid electrodes and screened for artifacts. Contaminated trials (as well as trials in which the pupil recording was marked as bad) were excluded.

In the PCA (conducted using the ERP PCA toolkit), 14 spatial factors were extracted from the subject ERPs (averaged for each stimulus type) and rotated using Promax without Kaiser normalization. Often, spatial PCA is followed by a temporal PCA. However, this method provides one temporal factor with a fixed latency for each ERP component and is therefore inappropriate to measure latency differences within a component (Dien, Spencer, & Donchin, 2004). Instead, we therefore picked the maximum factor score in the baseline-corrected "virtual ERPs" (spatial factor scores over time, Spencer et al., 1999) within time windows selected by visual inspection of the grand-average waveform (e.g., Brumback, Arbel, Donchin, & Goldman, 2012). Quantification of ERP components in individual trials, for which we used self-written MATLAB code, was analogous: The EEG was reduced to "virtual electrodes" (one for each spatial factor; see Goldstein et al., 2002), followed by peak picking. It is worth emphasizing that no new PCA was conducted on individual trials, but that the existing factor score coefficients were applied.

Pupil size. Pupil diameter was recorded from both eyes at 60 Hz using SmartEye Pro 5.8 software and two SmartEye Pro cameras installed below the screen. Using self-written MATLAB code, data points that were contaminated by artifacts were replaced by linear interpolation. Trials were excluded if more than 33% of the data points or 15 sequential data points were contaminated, or if the trial contained an artifact in the EEG. Next, we averaged across measurements from both eyes and applied a six-point moving average filter. We extracted the same trial types as for the ERPs using a time window of 500 ms before to 2,000 ms after stimulus onset. Mean diameter in the 500-ms prestimulus baseline was extracted and used to baseline-correct each trial, after which subject averages were calculated. Based on the waveform shape, we then extracted the maximum dilation (in mm),[4] and latency thereof (in ms) in a time window of 1,000 to 1,600 ms after stimulus onset, as well as the mean dilation in a second (return-to-baseline) time window (1,600–2,000 ms). Single-trial PDR quantification was analogous.

Trial rejection. We excluded trials from both analyses if either the ERP or the pupil size recording was marked as bad. This stringent rejection procedure resulted in the following trial numbers: frequents: 79–301, $M = 223.4$, $SD = 66.8$; infrequents: 17–58, $M = 43$, $SD = 12.3$; pictures: 4–29, $M = 19.6$, $SD = 7.1$.

Statistical Analysis

For statistical analysis, we used IBM SPSS software. Most analyses of the behavioral data and the factor scores (as measures of component amplitudes) used paired samples t tests or repeated measures analyses of variance (ANOVAs). Where the Mauchly test was significant, we report Greenhouse-Geisser-corrected p values. Post hoc tests used lower level ANOVAs or t tests with Bonferroni-corrected significance levels.

For the subsequent memory analysis, we averaged the physiological measures separately for trials that were, versus trials that were not, subsequently produced in the free recall test. Note that this analysis included only 19 participants, because one participant provided only two artifact-free trials for one of the ERP averages.

We examined covariance of physiological measures with RT by (a) analyzing subject averages for fast and slow responses using a median split, and (b) conducting single-trial correlation and regression analyses. The pattern of results was the same, so we report only the latter analysis. For this analysis, we standardized each physiological measurement as well as log RT by calculating z scores separately for each participant and measure. Trials with erroneous responses were not included. We then calculated Pearson's correlations between the physiological variables and RT (separately for each stimulus type). Variables that significantly correlated with RT were then included as predictors into linear multiple regression models using the forced entry method. We report standardized regression coefficients.[5]

Results

Behavioral Results

There were no significant differences in RTs and error rates between the three tasks or between blocks including versus not including pictures, so we collapsed across blocks. Table 2 provides a summary of the behavioral data.

Encoding. Error rates, $F(2,38) = 86$, $p < .01$, $\eta_p^2 = .82$, and RTs, $F(2,38) = 38.65$, $p < .01$, $\eta_p^2 = .67$, differed between stimulus types: Responses were less accurate and slower for infrequents than for frequents and pictures (all ts > 5.81), with no differences between the latter two (both ts < 1.8).

Recall rates. Participants recalled about 20 percent of the stimuli ($M = .2$, $SD = .05$). Recall rates differed between stimulus types, $F(2,38) = 16.26$, $p < .01$, $\eta_p^2 = .46$: Frequents were recalled with a lower probability than infrequents and pictures (both ts > 4.43), with no difference between the latter two ($t < 1.16$).

Recognition performance. Recognition responses were very accurate ($d' = 2.29$, $SD = .52$), and exhibited a liberal response

4. We also analyzed mean amplitudes for this time window. The patterns were analogous, so we here report only the maximum amplitude, which complements better the latency measure.

5. Note that, in additional analyses, we also included participant as random effect and word length (which indexes luminance) and word frequency as covariates into the regression model. The pattern of results was the same.

Table 2. *Means (SD) for Behavioral Data Measures*

		Frequents	Infrequents	Pictures
Encoding	Proportion correct	.96 (.04)	.76 (.1)	.97 (.05)
	RT (ms)	619 (73)	729 (99)	637 (86)
Recall	Proportion recalled	.16 (.06)	.29 (.11)	.33 (.15)
Recognition	Sensitivity (d')	1.68 (.26)	2.37 (.54)	3.05 (.73)
	Bias (c)	−.53 (.28)	−.25 (.26)	−.01 (.39)
	RT hit (ms)	795 (141)	815 (136)	674 (141)
	RT correct rejection (ms)	1173 (470)	1078 (320)	919 (280)

Note. In the RT data for the recognition phase, only correct responses given with high confidence are included. See text for details. RT = reaction time.

bias ($c = -0.37$, $SD = .24$). Both d', $F(2,38) = 61.32$, $p < .01$, $\eta_p^2 = .76$, and c, $F(1.53,29.14) = 18.07$, $p < .01$, $\eta_p^2 = .49$, differed between stimulus types: d' and c were smallest (indicating the lowest accuracy and the most liberal response bias) for frequents, intermediate for infrequents, and largest for pictures. Planned comparisons between all pairs reached significance after Bonferroni correction (all ts > 3.9), except for the bias difference between infrequents and pictures ($t < 2.2$).

Because not all participants provided responses of all types, we analyzed recognition RTs only to high confidence correct responses (i.e., *definitely old* responses to old items and *definitely new* responses to new items). The Response (hit/correct rejection) × Stimulus Type (frequent/infrequent/picture) ANOVA resulted in a main effect for stimulus type, $F(1.23,23.3) = 6.09$, $p < .05$, $\eta_p^2 = .24$, qualified by an interaction, $F(1.1,20.84) = 15.98$, $p < .01$, $\eta_p^2 = .46$. Correct rejections were faster for pictures than both infrequents and frequents, $F(1.17,22.17) = 8.85$, $p < .01$, $\eta_p^2 = .32$. For hits, differences were not significant, $F(2,38) = 2.38$, *ns*.

Summary. Infrequents (which required a switch to the infrequent response), but not pictures (requiring the frequent response) led to longer RTs and higher error rates than frequents. In turn, both infrequents and pictures exhibited an advantage over frequents in both the recall and recognition memory tests. Pictures enjoyed a further advantage over the infrequents in recognition performance.

Physiological Measures: Variance with Stimulus Type

ERPs. The grand-average ERPs elicited by infrequents and pictures exhibited larger positive amplitudes than for frequents (Figure 2A). The spatial PCA solution, which accounted for 91% of the total variance, suggested the elicitation of both a P300 and a novelty P3: Based on their spatial distributions and the virtual ERPs (Figure 2B), the frontocentrally distributed spatial factor (SF) 1 (variance accounted for: 32%) and the posterior SF 4 (9%) encompassed the novelty P3 and P300, respectively.

P300. Within SF 4, pictures elicited an earlier positivity than the other stimuli, so the time window for peak picking was set to 200–400 ms for pictures, but to 500–900 ms for frequents and infrequents. Quantified in this way, P300 amplitude differed between stimulus types, $F(1.49,28.25) = 7.19$, $p < .01$, $\eta_p^2 = .27$. Amplitude differences were not significant between infrequents and pictures ($t = 1.73$), but both differed from the frequents ($t > 2.84$).

Novelty P3. Novelty P3 amplitude, quantified as the maximum factor score of SF 1 in a time window of 400–800 ms, differed between stimulus types, $F(1.42,26.96) = 13.21$, $p < .01$, $\eta_p^2 = .41$. Amplitudes elicited by pictures and infrequents were larger than for frequents ($t > 3.48$), while at the corrected significance level the difference between infrequents and pictures was only a trend ($t = 2.45$).

PDR. The stimulus-elicited change in pupil size from the pretrial baseline was characterized by a dilation that peaked about 1,200–1,500 ms after the stimulus (Figure 2C). Quantified by the maximum between 1,000 and 1,600 ms, $F(2,38) = 15.24$, $p < .01$, $\eta_p^2 = .45$, as well as by the mean amplitude in the return-to-baseline time window (1,600–2,000 ms), $F(2,38) = 9.15$, $p < .01$, $\eta_p^2 = .33$, infrequents elicited a larger PDR than both frequents and pictures (all ts > 3.02) with no difference between the latter two ($t < .9$). Peak latency also differed between stimulus types, $F(2,38) = 4.26$, $p < .05$, $\eta_p^2 = .18$, with an earlier peak for frequents than infrequents ($t = 2.76$; all other ts < 1.8).

Supplementary analysis of luminance effects. Frequents and infrequents did not differ in luminance. Pictures extended the same visual angle as the longest words, but most pictures covered a larger area, resulting in lower luminance. Lower luminance increases pupil size (e.g., De Groot & Gebhard, 1952), so the smaller dilation for pictures compared to infrequents is unlikely to be due to luminance. Nevertheless, we contrasted pupil responses for pictures that were high versus low in luminance, as determined by a median split (Figure 3). There was a nonsignificant ($t < 1.5$) trend for pictures low in luminance to elicit larger PDRs. Therefore, luminance did not account for the PDR difference between pictures and infrequents.

Summary. The P300, the novelty P3, and the PDR differed in their sensitivity to stimulus type. Perhaps the most striking difference is the double dissociation between the novelty P3 and the PDR: Pictures elicited the largest novelty P3, while infrequents elicited the largest PDR.

Physiological Measures: Correlations with Behavior

The correlation analyses of the physiological measures with behavior focused on frequents and infrequents, because there were not enough picture trials for such analyses. Figures 4 and 5 show physiological responses split into fast and slow responses (using a within-subject median split), and by subsequent recall, respectively.

P300. For both frequents and infrequents, P300 latency was correlated with RT (Table 3), and continued to be a significant predictor of RT when the other variables had been accounted for (Table 4). It is worth noting that the amplitude difference between trials with

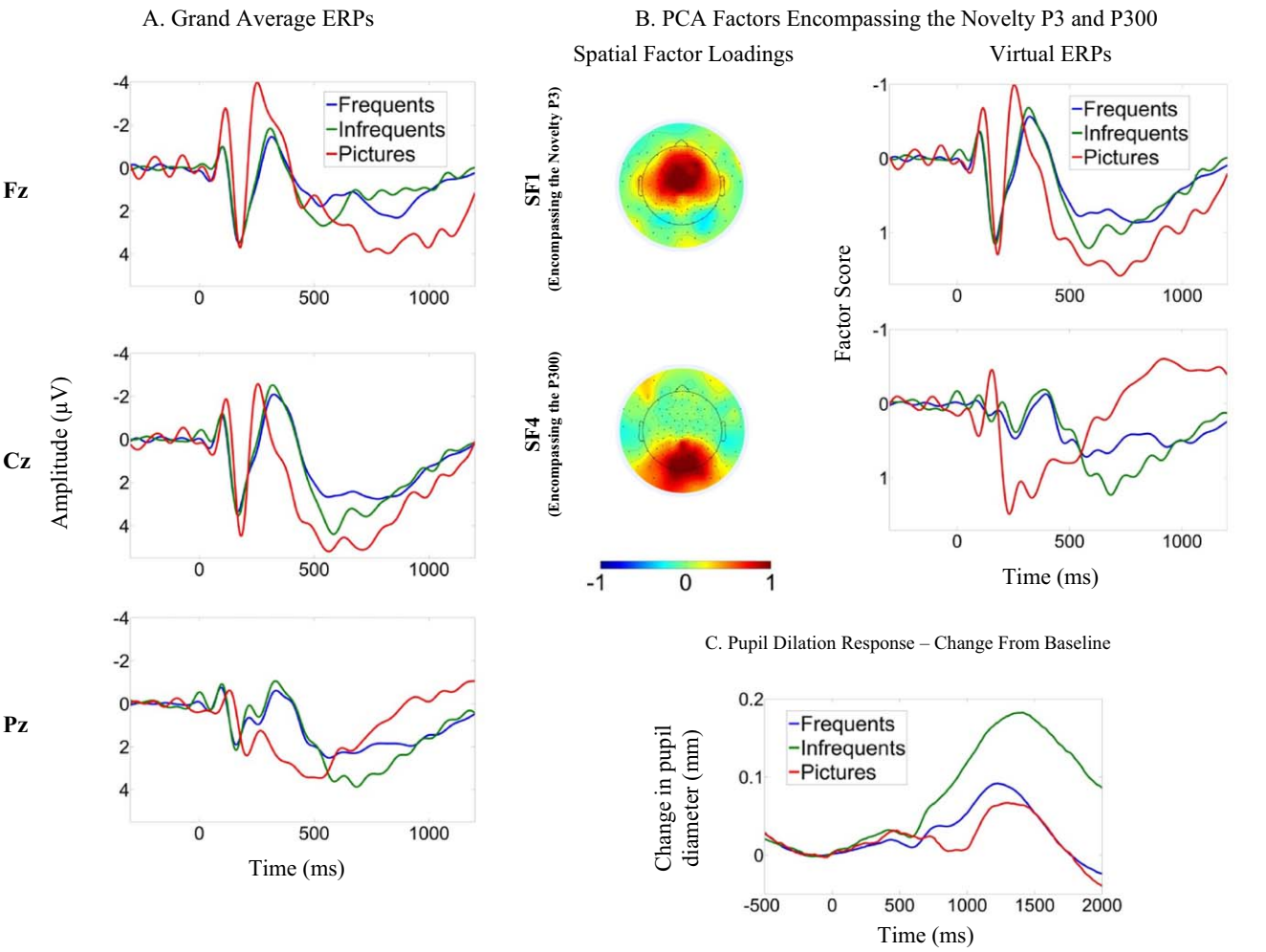

Figure 2. Response of ERP components and the PDR to stimulus types at encoding. A: Grand-average ERPs from frontal (Fz), central (Cz), and parietal (Pz) electrodes, time locked to the onset of frequents, infrequents, and pictures. B: Spatial PCA factors encompassing the novelty P3 and the P300, respectively. Displayed are spatial factor loadings (left panels), as well as factor scores plotted over time by stimulus type (virtual ERPs; right panels). C: Change of pupil diameter from baseline over time.

fast and slow responses, which is apparent in Figure 4, was neither statistically significant in the median split, nor in the correlation analysis. However, P300 amplitude was larger for subsequently recalled than not-recalled words, $F(1,18) = 12.95$, $p < .05$, $\eta_p^2 = .19$.

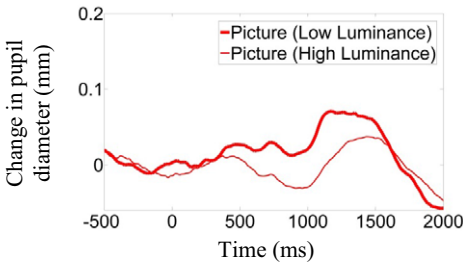

Figure 3. Change of pupil diameter from baseline, calculated for pictures lower in luminance versus pictures higher in luminance, as determined by a median split. The trend for low luminance pictures to elicit a larger PDR amplitude is not statistically significant.

Novelty P3. Larger novelty P3 amplitudes were consistently associated with faster responses to the frequents (Tables 3, 4); for the infrequents this negative correlation was nonsignificant (Table 3). For the infrequents, novelty P3 latency correlated with RT, but in the multiple regression model it became nonsignificant as a predictor of RT. Furthermore, the novelty P3 was larger for subsequently recalled compared to not-recalled stimuli, $F(1,19) = 6.45$, $p < .05$, $\eta_p^2 = .25$.

PDR. The PDR was prolonged and returned to baseline more slowly when RTs were long. Thus, PDR latency and PDR amplitude in the return-to-baseline time window were significantly correlated with RT (Table 3). For the infrequents, only PDR latency remained a significant predictor of RT in the multiple regression (Table 4). Larger PDRs were also elicited by subsequently recalled compared to not-recalled items [maximum: $F(1,18) = 17.55$, $p < .01$, $\eta_p^2 = .49$, mean in the return-to-baseline time window, $F(1,18) = 8.5$, $p < .01$, $\eta_p^2 = .31$].

Baseline pupil diameter. Pupil diameter during the pretrial baseline period predicted RT to frequents (Tables 3, 4). However, for

infrequents, baseline diameter did not correlate with RT. Baseline diameter did not exhibit a subsequent memory effect, $F(1,18) = .3$, *ns*.

Summary. P300 and PDR latency predicted unique portions of the variance in RT on the same trial. Furthermore, pretrial (baseline) pupil diameter, dilation in the return-to-baseline time window, and novelty P3 amplitude significantly predicted RT to frequents. Amplitudes of P300, novelty P3, and PDR all were larger for subsequently recalled than not-recalled stimuli.

Correlations Among Physiological Variables

Table 3 also shows the correlations between each pair of physiological variables. ERP amplitudes and latencies were relatively highly correlated with each other, and the same was true for PDR amplitude measures. However, P300 and novelty P3 amplitudes were uncorrelated with measures of PDR amplitude. Of further interest are the negative correlations between baseline pupil diameter and (a) PDR amplitude and dilation in the return-to-baseline time window, and (b) novelty P3 amplitude (which was statistically significant only for frequents). These negative correlations would be expected if baseline diameter is an index of tonic, and the PDR and novelty P3 reflect phasic activity in the LC.

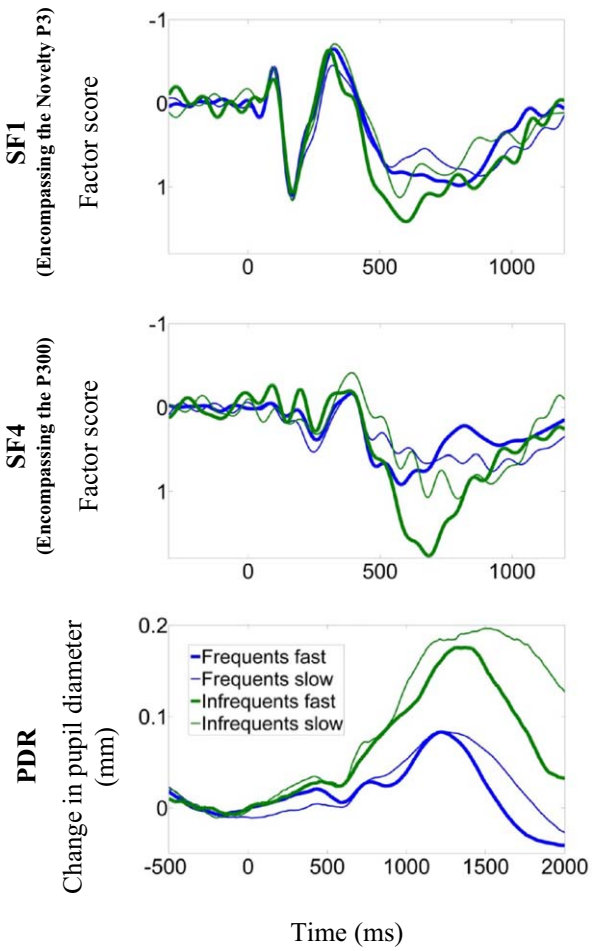

Figure 4. ERP and pupil responses for frequents and infrequents, split up by fast versus slow responses, as determined by a within-subject median split.

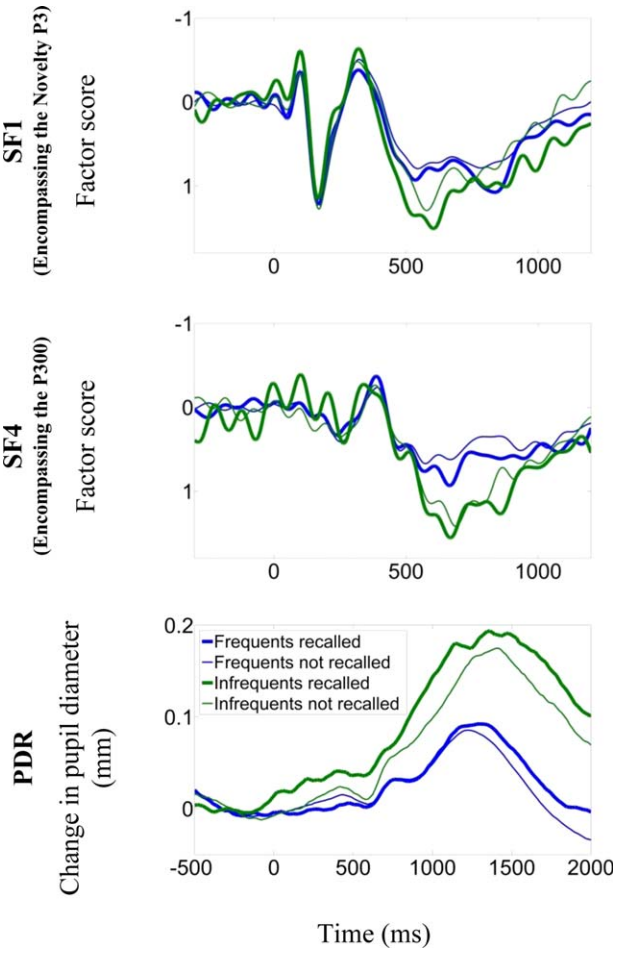

Figure 5. ERP and pupil responses for frequents and infrequents, split up by subsequently recalled versus subsequently not-recalled stimuli. [Correction added on 2 February 2015, after first online publication: the scale of the y-axis in the PDR panel and the figure legend have been amended.]

Discussion

We investigated the relationship between P300, novelty P3, and PDR by comparing their sensitivity to experimental manipulation and their within-subject correlations with RT and subsequent recall in a modified novelty oddball paradigm. The overarching question was whether the physiological responses are central and peripheral nervous system indicators of the same cognitive process (such as response facilitation), as suggested by hypotheses derived from a potential common innervation by the LC-NE system. If so, then they should (a) respond to experimental manipulations in the same way, (b) correlate similarly with RT and subsequent memory, and (c) be correlated with each other in individual trials. To foreshadow our discussion, our results do not appear to support the equivalence of each response's functional significance. We will first discuss the behavioral data from our paradigm, continue to summarize and outline implications of the findings on each physiological response individually, and finally discuss the relationship between the responses in the light of relevant functional theories.

Behavioral Data

Participants responded more slowly and less accurately to events of the infrequent category, but performed equally well for pictures as

Table 3. *Pearson's Correlation Coefficients for Each Pair of Physiological Variables*

		P300		Novelty P3		Pupil dilation response			
	1. RT	2. Ampl	3. Latency	4. Ampl	5. Latency	6. Ampl	7. Latency	8. Return to baseline	9. Baseline pupil diameter
1.	–	–	–	–	–	–	–	–	–
2.	−.01	–	–	–	–	–	–	–	–
	−.03								
3.	.07**	.10**	–	–	–	–	–	–	–
	.20**	.08*							
4.	−.07*	.48**	.03	–	–	–	–	–	–
	−.06	.54**	−.01						
5.	.01	.06**	.08**	.10**	–	–	–	–	–
	.17**	.06	.18**	.05					
6.	−.01	.00	.04*	−.01	−.01	–	–	–	–
	−.04	.03	.04	.04	.04				
7.	.14**	−.03*	.02	.01	.03	.11**	–	–	–
	.19**	−.05	.02	−.04	−.01	.13**			
8.	.08**	−.02	.01	−.02	−.01	.76**	.29**	–	–
	.10**	−.03	.05	−.04	.01	.79**	.35**		
9.	.08**	−.02	−.02	−.05**	−.01	−.35**	−.03*	−.43**	–
	.01	−.01	−.03	−.06	−.02	−.39**	−.07	−.43**	

Note. Correlation coefficients were calculated separately for frequents (first value in each cell) and infrequents (second value). For frequents, $df = 4282$; for infrequents, $df = 685$. RT = reaction time (at encoding); Ampl = amplitude.
**$p < .01$. *$p < .05$.

for frequents. Most likely, the frequent response is preplanned and must be inhibited while the alternative response is prepared when the infrequent category is presented, leading to response conflict only for the infrequent category. By contrast, both infrequents and pictures were more likely to be retrieved in the memory tests, which is in line with an abundance of data beginning with Von Restorff (1933). The pictures exhibited a further advantage over infrequents—an instance of the "picture superiority effect" on episodic memory (Shepard, 1967).

Table 4. *Multiple Regression Models on the Log Reaction Times at Encoding*

Variable	Standardized coefficient (β)	*SE*	*t*
A. Frequents			
P300			
Latency	.07**	.02	4.47
Novelty P3			
Amplitude	−.07**	.02	4.42
PDR			
Latency	.12**	.02	7.36
Mean return to baseline	.10**	.02	5.68
Baseline pupil diameter	.12**	.02	7.19
B. Infrequents			
P300			
Latency	.20**	.03	5.27
Novelty P3			
Latency	−.02	.03	.53
PDR			
Latency	.17**	.03	4.43
Mean return to baseline	.03	.03	.49

Note. Multiple linear regression analysis for trials of the frequent category ($R^2 = .04$**) and the infrequent category ($R^2 = .07$**). Only physiological variables that individually correlated with reaction time on the same trial were included in each model. *SE* = standard error.
**$p < .01$.

Variance of P300, Novelty P3, and PDR with Experimental Manipulation and Behavior

P300. The P300 was larger for infrequents than frequents—the typical P300 effect observed in oddball paradigms (Donchin, 1981). The pictures also elicited a parietal positivity that exhibited the same distribution as the P300 elicited by infrequents (note that we additionally confirmed this in PCA analyses conducted separately on the ERPs of each stimulus type). However, its latency was much shorter than for the infrequents, a difference ($\sim$ 400 ms) that was perhaps larger than might be expected solely because perceptual deviance (picture vs. word) can be detected more quickly than semantic deviance (infrequent vs. frequent), or because pictures are more quickly processed than words. Thus, while we cautiously interpret this positivity as an instance of the P300, future research should confirm its identity.

We replicated the finding that P300 amplitude was larger for subsequently recalled than not-recalled items (the P300 subsequent memory effect; Karis et al., 1984), supporting the idea that the cognitive process indexed by the P300 operates in interaction with episodic memory, as proposed by the context updating hypothesis. Participants' relatively high response accuracy at encoding (at least for the frequents) suggests that they behaved similarly to the condition in which response accuracy was emphasized in Kutas et al. (1977), so the correlation between P300 latency and response speed is a replication of those (and other) findings. However, based on the context updating hypothesis we also predicted that, when variance due to PDR latency (presumably as an index of response facilitation) had been statistically accounted for, P300 latency would not continue to explain a significant portion of the RT variance. Our results disconfirmed this prediction, so viewed in isolation our P300 results are not inconsistent with the idea that P300 is a direct link in the stimulus-response stream and cannot distinguish between the context updating hypothesis and the LC-NE theory of P300.

Novelty P3. In line with Spencer et al. (1999) and others, the frontocentral novelty P3 was largest for the pictures (which in our paradigm represented the perceptually salient novels), followed by infrequents, and finally frequents. This pattern notably paralleled subsequent memory performance, suggesting that the novelty P3 indexes processes that directly or indirectly aid episodic encoding. Novelty P3 amplitude was thus associated with (a) detection of (perceptual) deviance, (b) quicker behavioral responses (at least to the frequents), and (c) a higher probability of subsequent recall. The novelty P3 likely does not directly index response preparation or inhibition, as its latency did not account for unique variance in RT on the same trial. Rather, it may be an index of trial-by-trial variance in resource allocation (or reallocation) to novel stimuli (see also Friedman et al., 2001). Worth discussing in this context is a study by Wickens, Kramer, Vanasse, and Donchin (1983). In a dual-task paradigm, they varied the difficulty of a primary task to manipulate the extent to which resources were available for the secondary (oddball) task. P300 amplitude elicited in the secondary task was correlated with the amount of allocated resources. Since in such early studies dense electrode arrays were not available, and due to the spatiotemporal overlap between P300 and novelty P3, it is possible that the variance in scalp-recorded ERPs with resource allocation was, in fact, driven by variance in the novelty P3. It would be fruitful to explore this idea in future studies.

PDR. In sharp contrast to the novelty P3 and P300, the PDR was largest for infrequents and did not differ between pictures and frequents. The PDR therefore does not vary with expectancy violation per se, as infrequents and pictures were equally probable. Since only infrequents required the infrequent response, response conflict may be the crucial variable to elicit a large PDR. This is in line with results from studies using the Stroop task (e.g., Laeng et al., 2011; Siegle et al., 2004) or the Simon task (van Steenbergen & Band, 2013) and studies that varied the number of response alternatives (Richer & Beatty, 1987). Moreover, it is consistent with Kahneman's (1973) suggestion that pupil size indexes cognitive "effort." The individual-trial correlation between PDR latency (as well as the latency of the return of pupil diameter to baseline) and RT discovered in our study is also consistent with a role in facilitation of behavioral responding.

A possible alternative explanation is that the PDR indexes semantic deviance detection, because infrequents were also of a different semantic category than all other stimuli. However, in this case the infrequents should have also elicited a larger N400 (Kutas & Hillyard, 1980) than frequents and pictures, a pattern that was not apparent in our ERP data. Furthermore, in a prior study, the PDR negatively correlated with N400 amplitude (Kuipers & Thierry, 2011), speaking against the semantic deviance idea.

Relationship Between P300, Novelty P3, and PDR

P300 and novelty P3 amplitudes were uncorrelated with PDR amplitude. Together with their dissociable variance with stimulus type and their distinct covariance with behavioral measures (especially for novelty P3 vs. PDR), this provides evidence that these three physiological responses that are elicited in similar situations, nevertheless, play different functional roles. It is worth noting that the double dissociation between novelty P3 and PDR has also been observed, between studies, in go/no-go tasks: The no-go stimuli elicit a component that might be analogous to the novelty P3 (Pfefferbaum et al., 1985), while the go stimuli are the ones that elicit a larger PDR (Richer & Beatty, 1987).

Physiological responses and RT. P300 and PDR latency remained significant predictors of RT when the other responses had been accounted for. The predicted dissociation between the association of PDR and P300 latency with reaction time was therefore not obtained. The idea that the cognitive process indexed by the PDR may be integrated into the stimulus-response stream (as suggested by the LC-NE theory), while the P300 is not (as suggested by the context updating hypothesis) should, however, be further tested. That is, the fact that P300 and PDR latency accounted for unique portions of the RT variance could in theory be due to P300 latency being linked more closely to stimulus evaluation, while PDR is linked to responding, processes that both (perhaps relatively independently) affect RT. Independently manipulating stimulus processing and response demands, such as in the paradigm of McCarthy and Donchin (1981), would be ideal to test this.

Subsequent memory effects. Larger amplitudes of P300, novelty P3, and PDR were all associated with subsequently recalled compared to not-recalled stimuli. Further research is necessary to determine whether the circumstances in which each subsequent memory effect is pronounced exhibit perfect overlap or whether they are dissociable. For example, strategic encoding or retrieval processes may differentially affect which of the different responses will exhibit a subsequent memory effect.

The LC-NE theory of P300 and PDR: Correlations with baseline pupil diameter. Although we have thus far argued that the P300 and novelty P3 are functionally dissociable from the PDR, this does not necessarily exclude the possibility that all responses, at least in part, emerge from phasic LC innervation. That is, neurotransmitters and neural structures other than the LC-NE system likely contribute to the P300 (and novelty P3; Nieuwenhuis et al., 2011; Soltani & Knight, 2000), contributing to its functional significance. To explore whether each physiological response exhibits patterns consistent with the phasic LC response, we took baseline pupil diameter as an index of tonic LC activity (e.g., Gilzenrat et al., 2010), and investigated which of the physiological responses negatively correlates with baseline diameter (as would be expected for the phasic LC response). Similarly to PDR amplitudes, novelty P3 amplitude indeed exhibited this pattern. For the P300, however, a robust negative correlation with baseline diameter was not obtained.

Baseline pupil diameter and task performance. A final point that warrants discussion is that, although baseline (i.e., pretrial) pupil diameter correlated with RT, this was only true for the frequents, but not the infrequents. This is somewhat in conflict with, for example, Murphy et al.'s (2011) results, who found a correlation of baseline pupil diameter with RT to the infrequents in an auditory oddball task. Possible explanations for this difference in results include that (a) those authors used a much simpler oddball task in which participants had to distinguish between a high and a low tone, while in our paradigm a complex semantic judgment was performed; and (b) in their study, only the infrequents but not the frequents had to be responded to.

It is further worth noting that Murphy et al. (2011) reported an association between baseline diameter and performance that followed an inverted U-shape pattern. Additional analyses of our data suggested that the association was linear. Again, this might be explained by the differences in experimental design, and an interpretation of the different result should await further research.

Summary and Conclusions

Our results provide evidence that the functional significance of novelty P3, P300, and the PDR is dissociable. However, we did not find the predicted dissociation between P300 latency and PDR latency in their association with reaction time on the same trial. Future studies should further investigate this pattern and extend the present findings by employing paradigms that manipulate stimulus evaluation time and response preparation demands independently. Furthermore, additional studies should attempt to disentangle subsequent memory effects of each physiological response by carefully manipulating the demands on episodic encoding and retrieval. This way it would be possible to precisely determine under which circumstances P300, novelty P3, and PDR amplitude correlate with the probability of successful retrieval on a subsequent episodic memory test. Overall, such studies would contribute greatly to a more detailed theoretical framework for the functional significance of each physiological response.

References

Arbel, Y., Spencer, K., & Donchin, E. (2010). The N400 and the P300 are not all that independent. *Psychophysiology*, *48*, 861–875. doi: 10.1111/j.1469-8986.2010.01151.x

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450.

Aston-Jones, G., Rajkowski, J., & Cohen, J. (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biological Psychiatry*, *46*, 1309–1320.

Berridge, C. W., & Waterhouse, B. D. (2003). The locus coeruleus–noradrenergic system: Modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews*, *42*, 33–84.

Brumback, T., Arbel, Y., Donchin, E., & Goldman, M. S. (2012). Efficiency of responding to unexpected information varies with sex, age, and pubertal development in early adolescence. *Psychophysiology*, *49*, 1330–1339. doi: 10.1111/j.1469-8986.2012.01444.x

Butterfield, B., & Mangels, J. A. (2003). Neural correlates of error detection and correction in a semantic retrieval task. *Cognitive Brain Research*, *17*, 793–817. doi: 10.1016/S0926-6410(03)00203-9

Courchesne, E., Hillyard, S. A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology*, *39*, 131–143. doi: 10.1016/0013-4694(75)90003-6

Cycowicz, Y. M., & Friedman, D. (1999). The effect of intention to learn novel, environmental sounds on the novelty P3 and old/new recognition memory. *Biological Psychology*, *50*, 35–60. doi: 10.1016/s0301-0511(99)00004-6

Cycowicz, Y. M., & Friedman, D. (2004). The old switcheroo: When target environmental sounds elicit a novelty P3. *Clinical Neurophysiology*, *115*, 1359–1367.

De Groot, S. G., & Gebhard, J. W. (1952). Pupil size as determined by adapting luminance. *Journal of the Optical Society of America*, *42*, 492–495.

Dien, J. (2010). The ERP PCA toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods*, *187*, 138–145.

Dien, J., Spencer, K. M., & Donchin, E (2004). Parsing the late positive complex: Mental chronometry and the ERP components that inhabit the neighborhood of the P300. *Psychophysiology*, *41*, 665–678.

Donchin, E. (1981). Surprise! . . . Surprise? *Psychophysiology*, *18*, 493–513.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*, 357–374. doi: 10.1017/S0140525X00058027

Donchin, E., Heffley, E., Hillyard, S. A., Loveless, N., Maltzman, I., Öhman, A., . . . Siddle, D. (1984). Cognition and event-related potentials II. The orienting reflex and P300. *Annals of the New York Academy of Sciences*, *425*, 39–57. doi: 10.1111/j.1749-6632.1984.tb23522.x

Donchin, E., Spencer, K. M., & Dien, J. (1997). The varieties of deviant experience: ERP manifestations of deviance processors. In G. J. M. Boxtel & K. B. E. Bocker (Eds.), *Brain and behavior: Past, present and future*. Tilburg, The Netherlands: Tilburg University Press.

Duncan-Johnson, C. C., & Kopell, B. S. (1981). The Stroop effect: Brain potentials localize the source of interference. *Science*, *214*, 938–940.

Fabiani, M. (2006). Multiple electrophysiological indices of distinctiveness. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 85–119). Cambridge, MA: Oxford University Press.

Fabiani, M., & Donchin, E. (1995). Encoding processes and memory organization: A model of the Von Restorff effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 224–240.

Fabiani, M., Karis, D., & Donchin, E. (1986). P300 and recall in an incidential memory paradigm. *Psychophysiology*, *23*, 298–308.

Fabiani, M., Karis, D., & Donchin, E. (1990). Effects of mnemonic strategy manipulation in a Von Restorff paradigm. *Electroencephalography and Clinical Neurophysiology*, *75*, 22–35.

Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*: Boston, MA: Houghton Mifflin.

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience Biobehavioral Reviews*, *25*, 355–373. doi: 10.1016/s0149-7634(01)00019-7

Friedman, D., Hakerem, G., Sutton, S., & Fleiss, J. L. (1973). Effect of stimulus uncertainty on the pupillary dilation response and the vertex evoked potential. *Electroencephalography and Clinical Neurophysiology*, *34*, 475–484. doi: 10.1016/0013-4694(73)90065-5

Friedman, D., Simpson, G., & Hamberger, M. (1993). Age-related changes in scalp topography to novel and target stimuli. *Psychophysiology*, *30*, 383–396. doi: 10.1111/j.1469-8986.1993.tb02060.x

Gaeta, H., Friedman, D., & Hunt, G. (2003). Stimulus characteristics and task category dissociate the anterior and posterior aspects of the novelty P3. *Psychophysiology*, *40*, 198–208. doi: 10.1111/1469-8986.00022

Gilzenrat, M., Nieuwenhuis, S., Jepma, M., & Cohen, J. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience*, *10*, 252–269. doi: 10.3758/cabn.10.2.252

Goldinger, S. D., & Papesh, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science*, *21*, 90–95. doi: 10.1177/0963721412436811

Goldstein, A., Spencer, K. M., & Donchin, E. (2002). The influence of stimulus deviance and novelty on the p300 and novelty P3. *Psychophysiology*, *39*, 781–790. doi: 10.1111/1469-8986.3960781

Grider, R. C., & Malmberg, K. J. (2008). Discriminating between changes in bias and changes in accuracy for recognition memory of emotional stimuli. *Memory & Cognition*, *36*, 933–946.

Ilan, A. B., & Polich, J. (1999). P300 and response time from a manual Stroop task. *Clinical Neurophysiology*, *110*, 367–373.

Jepma, M., Deinum, J., Asplund, C. L., Rombouts, S. A. R. B., Tamsma, J. T., Tjeerdema, N., . . . Nieuwenhuis, S. (2011). Neurocognitive function in dopamine-beta-hydroxylase deficiency. *Neuropsychopharmacology*, *36*, 1608–1619. doi: 10.1038/npp.2011.42

Kafkas, A., & Montaldi, D. (2011). Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *Quarterly Journal of Experimental Psychology*, *64*, 1971–1989.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Kamp, S.-M., Brumback, T., & Donchin, E. (2013). The component structure of ERP subsequent memory effects in the Von Restorff paradigm and the word frequency effect in recall. *Psychophysiology*, *50*, 1079–1093.

Karis, D., Fabiani, M., & Donchin, E. (1984). "P300" and memory: Individual differences in the von Restorff effect. *Cognitive Psychology*, *16*, 177–216.

Kimmel, H. D. (1979). Prologue: What is the orienting reflex? In H. D. Kimmel, E. H. van Olst, & K. F. Orlebeke (Eds.), *The orienting reflex in humans*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kuipers, J. R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience*, *5*, 1–5. doi: 10.3389/fnhum.2011.00061

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

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, *197*, 792–795.

Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2011). Pupillary Stroop effects. *Cognitive Processing*, *12*, 13–21. doi: 10.1007/s10339-010-0370-z

Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology*, *21*, 171–186. doi: 10.1111/j.1469-8986.1984.tb00201.x

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science*, *211*, 77–80. doi: 10.1126/science.7444452

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior* (Vol. *960*). New York, NY: Holt.

Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus noradrenergic arousal function in humans. *Psychophysiology*, *48*, 1532–1543. doi: 10.1111/j.1469-8986.2011.01226.x

Naber, M., Frässle, S., Rutishauser, U., & Einhäuser, W. (2013). Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *Journal of Vision*, *13*.

Nieuwenhuis, S. (2011). Learning, the P3, and the locus coeruleus-norepinephrine system. In R. B. Mars, J. Sallet, M. F. S. Rushworth, & N. Yeung (Eds.), *Neural basis of motivational and cognitive control* (1st ed., pp. 209–222). Boston, MA: MIT Press.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, *131*, 510–532. doi: 10.1037/003-2909.131.4.510

Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*, *48*, 162–175. doi: 10.1111/j.1469-8986.2010.01057.x

Nuthmann, A., & Van Der Meer, E. (2005). Time's arrow and pupillary response. *Psychophysiology*, *42*, 306–317. doi: 10.1111/j.1469-8986.2005.00291.x

Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology*, *83*, 56–64.

Pfefferbaum, A., Ford, J. M., Weller, B. J., & Kopell, B. S. (1985). ERPs to response production and inhibition. *Electroencephalography and Clinical Neurophysiology*, *60*, 423–434.

Preuschoff, K., 't Hart, B. M., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience*, *5*, 1–12. doi: 10.3389/fnins.2011.00115

Richer, F., & Beatty, J. (1985). Pupillary dilations in movement preparation and execution. *Psychophysiology*, *22*, 204–207. doi: 10.1111/j.1469-8986.1985.tb01587.x

Richer, F., & Beatty, J. (1987). Contrasting effects of response uncertainty on the task-evoked pupillary response and reaction time. *Psychophysiology*, *24*, 258–262. doi: 10.1111/j.1469-8986.1987.tb00291.x

Rosenfeld, J. P., & Skogsberg, K. R. (2006). P300-based Stroop study with low probability and target stroop oddballs: The evidence still favors the response selection hypothesis. *International Journal of Psychophysiology*, *60*, 240–250. doi: 10.1016/j.ijpsycho.2005.05.010

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*, 156–163.

Siegle, G. J., Steinhauer, S. R., & Thase, M. E. (2004). Pupillary assessment and computational modeling of the Stroop task in depression. *International Journal of Psychophysiology*, *52*, 63–76.

Sokolov, E. N. (1963). Higher nervous functions: The orienting reflex. *Annual Review of Physiology*, *25*, 545–580.

Soltani, M., & Knight, R. T. (2000). Neural origins of the P300. *Critical Reviews in Neurobiology*, *14*, 199–224.

Spencer, K. M., Dien, J., & Donchin, E. (1999). A componential analysis of the ERP elicited by novel events using a dense electrode array. *Psychophysiology*, *36*, 409–414. doi: 10.1017/s0048577299981180

Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology*, *38*, 343–358.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*, 1187–1188.

van Steenbergen, H., & Band, G. P. (2013). Pupil dilation in the Simon task as a marker of conflict processing. *Frontiers in Human Neuroscience*, *7*. doi: 10.3389/fnhum.2013.00215

Verleger, R. (1997). On the utility of P3 latency as an index of mental chronometry. *Psychophysiology*, *34*, 131–156.

Võ, M. L. H., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., & Hutzler, F. (2008). The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophysiology*, *45*, 130–140.

Von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld [The effects of field formation in the trace field]. *Psychologische Forschung*, *18*, 299–342.

Wessel, J. R., Danielmeier, C., & Ullsperger, M. (2011). Error awareness revisited: Accumulation of multimodal evidence from central and autonomic nervous systems. *Journal of Cognitive Neuroscience*, *23*, 3021–3036. doi: 10.1162/jocn.2011.21635

Wickens, C., Kramer, A., Vanasse, L., & Donchin, E. (1983). Performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information-processing resources. *Science*, *221*, 1080–1082. doi: 10.1126/science.6879207

(Received May 21, 2014; Accepted September 29, 2014)