January 2013

The Psychophysiology of Novelty Processing: Do Brain Responses to Deviance Predict Recall, Recognition and Response Time?

Siri-Maria Kamp
University of South Florida, snkamp@mail.usf.edu

The Psychophysiology of Novelty Processing:

Do Brain Responses to Deviance Predict Recall, Recognition, and Response Time?

by

Siri-Maria Kamp

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Psychology
College of Arts and Sciences
University of South Florida

Major Professor: Emanuel Donchin, Ph.D.
Yael Arbel, Ph.D.
Kenneth J. Malmberg, Ph.D.
Geoffrey F. Potts, Ph.D.
Kristen Salomon, Ph.D

Date of Approval:
July 11, 2013

Keywords: Event-related potentials, pupillometry, P300, Novelty P3, oddball paradigm

Table of Contents

List of Tables

List of Figures

Abstract

Events that violate expectations are biologically significant and accordingly elicit various physiological responses. We investigated the functional relationship between three of these responses: the P300, the Novelty P3 and the pupil dilation response (PDR), with a particular focus on their co-variance with reaction time and measures of subsequent memory. In a modified Novelty P3 oddball paradigm, participants semantically categorized a sequence of stimuli including (1) words of a frequent category, (2) words of an infrequent category (14% of the trials) and (3) pictures of the frequent category (14% of the trials). The Novelty P3 oddball task was followed by a recall- and a recognition test. Larger amplitudes of the P300, identified by a spatial principal component analysis (PCA), were associated with enhanced subsequent recall as well as faster reaction times during the recognition test, suggesting a close relationship between the cognitive process indexed by the P300 and memory encoding. The PDR was larger for infrequents (which required a response switch) than both frequents and pictures (which did not require a switch). Furthermore, its latency was correlated with reaction time on the same trial and with reaction time on the immediately following trial. There was only weak evidence for a correlation with subsequent memory, suggesting that the cognitive process associated with the PDR might be a direct link in the stimulus-response stream. Larger Novelty P3 amplitudes were associated with both faster reaction times on the same trial and stronger memory traces, suggesting that its amplitude might index resource allocation. These findings suggest that each of the physiological responses

carries a distinct functional significance in detecting, processing, or responding to novel events, and we discuss the findings in the light of the prevalent theories of the functional significance of each response.

Introduction

The nervous system of humans and other animals preferentially attends to, detects, and mnemonically encodes novel, deviant or surprising events (e.g., Ranganath & Rainer, 2003). If an event is predictable from information available prior to the event's occurrence, usually no adjustment of the behavioral program is required. However, if a highly unexpected event occurs, it is adaptive to quickly prepare for action and to encode this event into long-term memory so that future behavior can be adjusted accordingly. In line with the biological significance of unexpected events, various physiological reactions occur in the human body when novelty is encountered. The present study focuses on a subset of these physiological processes and examines their functional significance in the processing of novel events, specifically the extent to with they index short-term and long-term behavioral responding to the novel event.

Novel, non-noxious stimuli elicit a set of autonomic responses, which habituate when the same stimulus is repeatedly encountered: The "orienting reflex" (see, for example Barry, 2009; Kimmel, 1979; Sokolov, 1963), including a temporary dilation of the pupil (the "pupil dilation response"; PDR). Stimulus deviance also elicits a number of cortical responses, as for example measured by event-related potentials (ERPs). For example, the Mismatch Negativity, N2, Novelty P3, P3a, P300, N400, and slow waves, are all invoked in response to stimulus deviance in one way or another (Donchin, Spencer, & Dien, 1997; Fabiani, 2006). The Novelty P3 and the P300 are of interest for the present study because they are sensitive to similar stimulus- and task parameters as the autonomic orienting

reflex in general (Donchin et al., 1984), and the PDR in specific (Nieuwenhuis, De Geus, & Aston-Jones, 2011).

Thus, the brain contains highly specialized areas that are involved in specific aspects of information processing, such as language production (Broca's area) or emotional reactivity (e.g. amygdala). At the same time, brain areas do not work in isolation, but every complex cognitive process engages an entire neural network. Therefore, each physiological response that is measurable when novelty is encountered may reflect specialized, unique cognitive processes (such as immediate responding vs. learning), or alternatively, the responses as a whole may reflect a unitary, wholistic process invoked to process, and respond to, novelty (as appears to be suggested by early reports of the orienting reflex, e.g. Sokolov, 1963). The present paper hypothesizes that while Novelty P3, P300 and PDR all are evoked by novel events, each physiological process reflects a separate function that is either directly integrated into the stimulus-response stream, such as perceptual sensitization or response adjustments, or "strategic" functions invoked in parallel to the stimulus-response stream, such as episodic memory encoding. The idea that different responses elicited by novelty reflect different functions is in line with proposals that "physiological processes characterized as signs of orienting may actually have taken place in very different poststimulus phases and that these processes may have reflected very different aspects of stimuli and organismic functioning" (Näätänen, 1978, p. 63).

There is some agreement in the literature that the collection of autonomic responses to novelty (e.g., Näätänen, 1978), including the PDR (Nieuwenhuis et al., 2011), reflect the sensory processing of, or facilitation of the immediate response to, novel events,

suggesting that these responses are directly integrated into the stimulus-response stream. In contrast, the most prominent theory of the P300, the *context updating hypothesis* (Donchin, 1981; Donchin & Coles, 1988), proposes that the P300 reflects processes that occur *in parallel* to the stimulus-response stream, such as strategic adjustments of *future* behavior, including episodic memory encoding. In combination with ideas from the orienting reflex literature, the context updating hypothesis would imply that the P300 and the PDR index separate, unique cognitive functions. However, this idea has been recently challenged by Nieuwenhuis and colleagues (Nieuwenhuis, Aston-Jones, & Cohen, 2005; Nieuwenhuis et al., 2011), who have suggested that the P300 indexes the optimization of action in response to deviant events. This would suggest that P300 and PDR index analogous processes related to immediate responding.

The two competing theories of P300 function are derived from different vantage points: While the context updating hypothesis is based on a study of the eliciting conditions and consequences of the P300, the latter theory has been derived from the putative physiological origin of the P300 in noradrenergic cortical input from the locus coeruleus (LC) of the midbrain (Nieuwenhuis, Aston-Jones, et al., 2005). Indeed, the quest of identifying the functional significance of brain activity is often approached with one of these two strategies: On the one hand, identifying eliciting conditions and consequences of the ERP component; and on the other hand identifying the anatomical origin(s) and then utilizing prior knowledge about the neural substrates to assess its function (Donchin et al., 1997). When conclusions derived from both vantage points converge, this can dramatically strengthen theoretical accounts; however, in the present case a conflict arises that raises the question whether or not the P300 (as well as the

Novelty P3) and the PDR reflect central- and peripheral nervous system analogues of the same function.

To examine this issue, the present study simultaneously investigates the functional significance of three physiological responses elicited by deviancy: The P300, the Novelty P3, and the PDR. The hypothesis is that although the responses share *antecedent conditions*, they might index different psychological *functions*: The PDR may be related to the facilitation of the adjustment of sensory- or motor processes required for immediate responding, while the P300 may index learning. Therefore, PDR should be correlated with reaction time on the same or subsequent trials, while P300 should predict memory performance.

The next chapter will provide a more detailed review of each physiological response of interest, including theories about its function as well as prior findings on its relationship to reaction time and episodic memory.

Background

The P300

The P300 was discovered in 1965 (Sutton, Braren, Zubin, & John, 1965) and has since become one of the most widely studied ERP components. It manifests as a positive deflection with a parietal maximum and peaks between 300 and 700ms after stimulus onset.

Basic Characteristics of the P300. The typical experiment known to elicit a P300 is the *oddball paradigm*, which consists of a sequence of stimuli that can be classified into one of two categories, of which one occurs rarely and one frequently. When the subject actively classifies the stimuli according to the two categories, the infrequent category elicits a P300. Other experiments that elicit a P300 all share the characteristic that the eliciting stimulus is rare, unexpected, or very salient, as well as task-relevant (for a review, see Donchin, 1981).

P300 amplitude is inversely correlated to the subjectively perceived stimulus probability within the sequence (Duncan-Johnson & Donchin, 1977; K. C. Squires, Wickens, Squires, & Donchin, 1976) and increases with the length of the inter-stimulus interval: for very long intervals even the frequent category can elicit a P300 (Polich, 1990). Furthermore, the P300 is larger when the inter-stimulus interval is fixed than when each event occurs after a random interval (Schwartze, Rothermich, Schmidt-Kassow, & Kotz, 2011). When the oddball task is a *secondary* task, P300 amplitude is inversely correlated to primary task difficulty, suggesting that its amplitude can be used

as an indicator of the "attentional resources" allocated to the primary task (Wickens, Kramer, Vanasse, & Donchin, 1983).

Different techniques have pointed to some likely candidate brain regions as generators of the P300, including the hippocampus (Axmacher et al., 2010; although the extent to which hippocampal activity can directly contribute to scalp-recorded EEG activity is controversial; e.g., Fernández et al., 1999) and the temporo-parietal junction (e.g., Knight, Scabini, Woods, & Clayworth, 1989). There is converging evidence from different techniques for each of these generators, but most likely a network of multiple brain regions including subcortical and cortical regions in the parietal, temporal, and occipital lobe, rather than an individual source, generates the scalp-recorded P300 (for reviews, seeKnight & Scabini, 1998; Linden, 2005).

Interestingly, the P300 is sensitive to similar experimental manipulations as are norepinephric (NE) neurons located in the locus coeruleus (LC) of the brain stem. Since ERPs reflect post-synaptic activity (as opposed to action potentials), and since the LC neurons broadly project to cortical areas including those that have been implicated in P300 generation, the P300 may be the consequence of NE emission by the LC towards these brain areas (Nieuwenhuis, Aston-Jones, et al., 2005; Nieuwenhuis et al., 2011).

The Context Updating Hypothesis of the P300. The *context updating hypothesis* (Donchin, 1981; Donchin & Coles, 1988) is the most influential theory on the functional significance of the P300 to date (Polich & Kok, 1995). The theory builds upon the model of information processing proposed by Miller, Galanter and Pribram (1960) in that it assumes that an individual maintains a mental schema of all presently goal relevant information. If new information conflicts with expectations derived from this schema, an

"updating" process is elicited through which this new information is incorporated into the schema. According to the context updating hypothesis, the P300 indexes the updating process and is thus proposed to reflect strategic processes affecting *future* behaviors, rather than an immediate behavioral reaction to the stimulus. In other words, the cognitive process associated with the P300 is not a direct link in the sequence between the perception of the stimulus and the behavioral response, but is part of a parallel processing stream that maintains and modifies the schema.

The schema cannot be updated, nor can an accurate behavioral reaction be executed, before the stimulus has been evaluated and classified as deviant. Therefore, both P300 latency and response time should both be longer when stimulus classification takes a longer time. However, the model predicts that the relationship between P300 latency and reaction time is not a necessary one, and that under some circumstances the latency of the P300 and reaction time are dissociable. At the same time, since context updating occurs in interaction with long-term memory, the theory predicts that P300 amplitude, indexing the strength of the updating process, will be correlated with the probability of later remembering the respective event (Donchin, 1981).

The LC-NE Theory of the P300. According to Nieuwenhuis and colleagues (2005), lesion studies, pharmacological studies and the functional connectivity of LC neurons with cortical regions suggest that the P300 reflects the synchronized norepinephric (NE) input to multiple cortical regions from neurons of the locus coeruleus (LC). The context updating hypothesis does not assume any specific P300 generators, so the physiological basis of the P300 within the LC-NE system does not contradict the context updating

hypothesis *per se*. However, the conclusions about P300 function that have been drawn from this theory are in conflict with the context updating hypothesis.

Thus, the response characteristics of LC neurons consist of a *tonic* and a *phasic* signal. The tonic signal indexes the degree to which the subject is engaged in goal-directed vs. exploratory behavior (or, more generally, their state of arousal). The phasic signal is time-locked to task-relevant events, and is enhanced for infrequent events and novels in oddball paradigms (cf. Nieuwenhuis, Gilzenrat, Holmes, & Cohen, 2005). The theory proposes that the P300 emerges from this phasic signal. There is much evidence, as reviewed in more detail in the chapter on the pupil dilation response, for a relationship of LC activity to behavioral responding. Therefore, the phasic NE input to the cortex, and consequently the P300, may "facilitate responding" to motivationally relevant stimuli (Nieuwenhuis, Aston-Jones, et al., 2005). Apparently in conflict to this idea, P300 latency is often longer than the reaction time, suggesting that the P300 cannot be a direct link in the stimulus-response stream. However, the Nieuwenhuis and colleagues have argued that this inconsistency may be artificial because P300 latency is typically defined as the time point of its *maximum*, when it might be more appropriate to define P300 latency by its *onset*.

There are some difficulties with Nieuwenhuis' review of the previous P300 literature. For example, their claim that P300 is elicited only in situations where a response is required is not supported by prior research (Duncan-Johnson & Donchin, 1977). Furthermore, Nieuwenhuis et al. (2005) conceptualize the P300 and the Novelty P3 as "sub-components" of the P300 as if they were not functionally distinct components. As will be reviewed in the next section, this does not accurately reflect the literatures on

P300 and Novelty P3 (Dien, Spencer, & Donchin, 2003; Spencer, Dien, & Donchin, 1999). Nevertheless, the functional similarities between LC activity and the P300 and their physiological connections are striking and taking these into consideration can augment theories of its functional significance.

The LC-NE theory differs from the context updating hypothesis in a fundamental way: The LC-NE theory places the cognitive processes associated with the P300 directly within the stimulus-response stream, while the context updating theory does not. To accommodate findings that implicate a relationship of LC activity to memory formation, recently Nieuwenhuis (2011) revised his theory, attributing cognitive processes associated with *both* immediate action and learning to the P300. This revised theory is much broader than, and therefore still not identical to the context updating hypothesis.

The context updating hypothesis and the LC-NE theory make different predictions about the relationship between P300, reaction time and memory. The next two sections review prior studies addressing these relationships.

P300 and Reaction Time. Both P300 latency and reaction time are sensitive to the time it takes to evaluate the stimulus: Both are longer for more difficult categorization tasks (e.g., Kutas, McCarthy, & Donchin, 1977). However, if the cognitive process indexed by the P300 is related to response preparation or execution, P300 latency should be more strongly associated with the behavioral response following the stimulus than to the stimulus onset itself.

In one of the first studies that investigated the relationship between P300 and reaction time, Kutas et al. (1977) applied a within-subject, trial-by-trial analysis of both measures. When participants were asked to respond *as accurately as possible*, P300 latency was

strongly correlated with reaction time. In contrast, when response *speed* was emphasized over accuracy, P300 latency and reaction time were uncorrelated. This was interpreted such that P300 latency depends on stimulus evaluation time, and when the behavioral response is also based on accurate stimulus evaluation, a correlation between the two measures is observed. However, speeded behavioral responses may be executed before a full evaluation of the stimulus, but since P300 amplitude still depends on stimulus evaluation time, the two measures are decoupled.

McCarthy and Donchin (1981) presented participants with stimuli that were (1) either embedded in a matrix of visual noise (difficult discrimination) or not embedded in noise (easy discrimination), and that (2) required a response that was congruent with the stimulus (e.g., a right hand response to the word "right"), or incongruent with the stimulus. Both reaction time and P300 latency increased with discrimination difficulty. However, while response incongruency led to longer response times, it left P300 latency unaffected. Magliero, Bashore, Coles and Donchin (1984) replicated this pattern and demonstrated a parametric relationship between P300 latency and discrimination difficulty degree of noise. Together with the study by Kutas et al., these findings suggest that P300 latency is more closely related to stimulus evaluation than the behavioral response. It is worth noting that in a similar paradigm, Smulders, Kenemans, Schmidt and Kok (1999) reported that stimulus-response incompatibility *did* increase P300 latency. However, in this study stimulus-response compatibility was manipulated between blocks and may therefore have been confounded with task-related factors. Furthermore, even in Smulders et al.'s findings, P300 latency varied more strongly with stimulus discrimination difficulty than with response incompatibility.

P300 latency and reaction time have also been dissociated in the Stroop task. For example, Ila and Polich (1999) asked participants to classify with a button press the font color of color words that were either congruent or incongruent with the font colors, or the font color of nonwords. Reaction times were shorter for congruent, and larger for incongruent, compared to neutral trials. However, P300 latency was unaffected by congruency.

In sum, P300 latency and reaction time often co-vary in situations in which reaction time is mostly determined by stimulus evaluation processes (Friedman, 1984; Holm, Ranta-aho, Sallinen, Karjalainen, & Müller, 2006; Leuthold & Sommer, 1993). However, when additional complexities affect response initiation, or when responses are based on incomplete stimulus evaluation, P300 latency and reaction time are often uncorrelated. The decoupling of P300 from reaction time suggests that the cognitive process associated with the P300 is not located directly within the stimulus-response stream, as suggested by the context updating hypothesis. However, Nieuwenhuis et al. (2011) argue a P300 is not *necessary* to elicit a behavioral response; rather, the cortical NE release that leads to the P300 is neuromodulatory. Therefore, *when* a P300 is elicited and occurs before the response, the associated cognitive process facilitates responding and leads to better behavioral performance. They support this suggestion with data on the relationship between P300 *amplitude* and task performance.

For example, P300 amplitude is positively correlated with the probability that participants detect the visual target that elicited the P300 (e.g., Rolke, Heil, Streb, & Hennighausen, 2001). P300 amplitude has also been reported to inversely correlate with reaction time on the same trial (Friedman, 1984). However, the positive relationship

between P300 amplitude and performance is not unequivocal: When a stimulus predicts the occurrence of a second stimulus at a certain probability, lower probabilities are associated with larger P300 amplitudes, but also with *longer* reaction times and *lower* accuracy (Duncan-Johnson & Donchin, 1982). In summary, the controversy of the relationship between P300 and reaction time (and task performance more generally) is, to date, not fully resolved.

P300 and Memory Encoding. In the "Von Restorff paradigm", named after its inventor (Von Restorff, 1933), a study list contains one item that is *distinctive,* for example due to its font size or color, while all other study items are identical in this feature. This "isolate" is more likely to be freely recalled than the other items – the "Von Restorff effect". Due to their infrequent, task-relevant nature, it is not surprising that the isolates also elicit a P300. The P300 is not an all-or-nothing response, and varies from one trial to the next, and the context updating hypothesis predicts that this variance in P300 amplitude is correlated with the variance in recall (Donchin, 1981).

In the first study that tested this hypothesis (Karis, Fabiani, & Donchin, 1984), study lists in the Von Restorff paradigm were followed by immediate free recall tests. Isolates in a deviant font size exhibited the typical recall advantage and elicited a P300, which was larger for subsequently recalled, compared to forgotten words. However, this P300 "subsequent memory effect" was only observed for participants who used rote rehearsal at encoding. For participants who used elaborative strategies (as well as for non-distinctive items), the *frontal positive slow wave* was correlated with recall (Karis et al., 1984). This pattern was replicated in an experiment in which participants were *instructed* to use specific encoding strategies (Fabiani, Karis, & Donchin, 1990). In a third study,

participants *incidentally* encoded names in an oddball paradigm (Fabiani, Karis, & Donchin, 1986), followed by a surprise recall test. Again, a P300 subsequent memory effect was observed, and since participants are unlikely to elaborate when encoding is incidental, this is in line with the idea that when elaborative strategies are *not* used, P300 amplitude is correlated with recall.

Finally, three studies manipulated the *manner* in which a word was isolated. Under item-based encoding tasks, the P300 subsequent memory effect occurred for words in a larger font size as well as isolates of a distinct semantic category (Fabiani & Donchin, 1995) and low frequency words embedded in a list of high frequency words (Kamp, Brumback, & Donchin, in press). However, for isolates with a frame drawn around them, the P300 was not correlated with recall, even under rote encoding strategies. Instead, similarly to non-distinctive words, the frontal slow wave showed a subsequent memory effect (Otten & Donchin, 2000). Thus, the correlation between P300 amplitude and recall occurs only if the isolating feature is *integral* to the study item.

Other lab groups have also used the subsequent memory paradigm on ERPs (for a review, see Paller & Wagner, 2002). Many of these studies did not identify the specific ERP components whose amplitudes were correlated with later memory, but rather termed any difference between the later remembered and later forgotten ERPs "difference due to memory" (Paller & Wagner, 2002). However, in line with the studies reviewed above, most subsequent memory studies have found more positive-going ERPs for later recalled, compared to not recalled words, particularly in parietal regions (e.g., Azizian & Polich, 2007; Kim, Vallesi, Picton, & Tulving, 2009; Paller, Kutas, & Mayes, 1987; Voss & Paller, 2009; Wiswede, Rüsseler, & Münte, 2007).

Nieuwenhuis et al. (2005) propose that due to the correlational nature of these patterns, the P300 subsequent memory effect could be explained by the influence of third variables on both P300 and memory encoding, such as a "time-locked heightening in selective attention". The present study will help answer this question, because concurrent measurement of multiple ERP components and the PDR will enable us to distinguish between a general heightening in attentional resources (as indicated by an increase in all physiological responses) and a more direct relationship between P300 and memory processes.

The Novelty P3

The second ERP component of relevance for the present study is the Novelty P3, an ERP component that is often categorized into the "P3 family" due to its temporal overlap with the P300 (Polich & Kok, 1995). However, the Novelty P3 is a distinct component that is spatially and functionally dissociable from the P300 (Spencer et al., 1999).

Eliciting Conditions. The Novelty P3 is a positive-going ERP component elicited by task-irrelevant, novel (as well as typically perceptually salient) stimuli. Courchesne et al. (1975) first reported the elicitation of this component in a modified visual oddball task that, among frequent "standard" stimuli (the digit 2) and infrequent "target" stimuli (the digit 4) also included the infrequent presentation of novel task-irrelevant stimuli, i.e., line drawings of unfamiliar objects. The novels elicited a fronto-central positivity which is now known as the Novelty P3 (Courchesne, Hillyard, & Galambos, 1975). Subsequent studies revealed that the Novelty P3 is also elicited by auditory novels, such as environmental sounds embedded in an oddball sequence of two simple tones (Friedman, Simpson, & Hamberger, 1993). Novel stimuli also elicit a P300, but the Novelty P3 and

the P300 are separate components with distinct spatial distributions (Spencer et al., 1999).

In the original paradigm, the novels were task-irrelevant. However, there is some evidence that task-irrelevance is not necessary for the elicitation of a Novelty P3. For example, when the participant is instructed to memorize the infrequently presented novel sounds for a later test (Cycowicz & Friedman, 1999), or when a button press is required to the novels (e.g., Cycowicz & Friedman, 2004; Gaeta, Friedman, & Hunt, 2003), a Novelty P3 is still elicited. Therefore, it appears that perceptual deviance is more crucial than task-irrelevance to elicit a Novelty P3.

In the same year in which the original Novelty P3 paper was published, another group reported a morphologically similar positivity, the P3a, elicited by infrequent (non-novel) events in oddball tasks in which the participant ignores the stimuli (N. K. Squires, Squires, & Hillyard, 1975). Simons, Graham, Miles and Chen (2001) showed that the two components are indistinguishable in their spatial and temporal distributions. In line with this idea, Spencer et al. (1999) demonstrated that in a Novelty P3 oddball paradigm, infrequent target stimuli also elicit a P3a/Novelty P3, but with a smaller amplitude than for the novels. Therefore, in the present paper the term "Novelty P3" will refer both to fronto-central positivities elicited by task-irrelevant novel stimuli and by infrequent, non-novel stimuli.

Functional Significance. In the original study by Courchesne et al. (1975), the authors suggested that the Novelty P3 reflects an "orienting response" to salient, unexpected, and task-irrelevant stimuli. This idea is still prevalent among researchers studying the Novelty P3 (for a review, see Friedman, Cycowicz, & Gaeta, 2001). Key

observations that support the similarity of the Novelty P3 to the "orienting reflex" as first described by Sokolov (based on prior work by Pavlov; e.g., Sokolov, 1963) are the habituation with repeated stimulus presentations and its sensitivity to stimulus saliency (Courchesne et al., 1975; Rushby & Barry, 2009). However, for the purpose of describing its functional significance, placing the Novelty P3 within the orienting reflex framework is not very useful. That is, the precise function of the orienting reflex is controversial, and different autonomic responses placed within the framework may serve different functions (see next chapter for details).

Another hypothesis is that the Novelty P3 indexes response inhibition (e.g., Goldstein, Spencer, & Donchin, 2002). That is, typical Novelty P3 paradigms require a response to all stimuli *but* the novels, so that a possibly pre-programmed response must be inhibited. Further evidence for this idea comes from the finding that a fronto-central, and therefore morphologically very similar, positivity is elicited by "No-Go" stimuli in a "Go/No-Go" paradigm (e.g., Pfefferbaum, Ford, Weller, & Kopell, 1985). If the two positivities were indeed instances of the same ERP component, this would be strong evidence for the response inhibition hypothesis. However, the response inhibition idea is inconsistent with the findings reviewed above that adding a response requirement to the novels does not abolish the elicitation of a Novelty P3.

Since studies on the relationship between Novelty P3 and behavior within subjects are rare, in discussing its functional significance another morphologically similar and perhaps functionally analogous ERP component is worth taking into consideration. Thus, when a participant commits a behavioral error that they can detect independently, or when a participant receives informative feedback that an error has been committed, an

error-related negativity (ERN), along with an error positivity (Pe) is elicited (Gehring, Goss, Coles, Meyer, & Donchin, 1993; Miltner, Braun, & Coles, 1997). The scalp-recorded Pe consists of a parietal P300 as well as a frontally distributed positivity with a Novelty P3-like morphology (Arbel & Donchin, 2009). The eliciting conditions can be considered similar to the Novelty P3 in that errors are infrequent, most likely salient, and thus "novel" events.

The relationship between the Novelty P3 (and Pe) to behavioral responding and subsequent memory has not been extensively studied within participants. In the typical Novelty P3 oddball paradigm, no response is required to the novels, and memory is not typically tested. However, there is some indirect evidence for a role of the Novelty P3 in response adjustments after a novel, unexpected event occurs: In a reaction time experiment, the occurrence of an infrequent, task-irrelevant tone slowed down the response to the immediately following stimulus (Notebaert et al., 2009). Although no ERPs were recorded, it is likely that the task-irrelevant, infrequent tone elicited a Novelty P3, and it can be speculated that the Novelty P3 amplitude might correlate with the reaction time to the subsequent trial.

Notebaert et al. (2009) also manipulated the probability of participants making erroneous responses. "Post-error slowing" – the extent to which the response to the trial after an erroneous response is slowed down (Rabbitt, 1969) – was only observed when errors were less likely than correct responses. When correct responses were less likely than errors, they reported slower reaction times following correct responses. It appears therefore that it is not the error *per se,* but rather the occurrence of an unexpected event that causes post-error slowing (Notebaert et al., 2009), indirectly supporting the idea that

Pe/Novelty P3 may index response adaptation due to the occurrence of a novel event. In line with this idea, the amplitude of the Pe has been reported to correlate with post-error slowing (Hajcak, McDonald, & Simons, 2003), although this pattern is not consistent within the literature (for a review see Overbeek, Nieuwenhuis, & Ridderinkhof, 2005).

Few studies have investigated the relationship between the Novelty P3 and memory. Fabiani and Friedman (1995) found that older adults exhibited both smaller Novelty P3 (and P300) amplitudes and reduced recognition memory accuracy for the novels. However, a subsequent memory analysis within participants was not conducted. In a visual Novelty P3 oddball paradigm, Cycowitz and Friedman (1999) found that neither under incidental, nor under intentional encoding instructions a correlation between Novelty P3 and the probability of successful subsequent recognition of the novels was observed. The P300, by contrast, showed a subsequent memory effect for the first novel item when it was intentionally encoded.

While Cycowicz and Friedman's (1999) study is the only one to investigate subsequent memory effects in a typical Novelty P3 oddball paradigm, other studies with less typical paradigms have reported subsequent memory effects within the Novelty P3. Using the Von Restorff paradigm, Kamp, Brumback and Donchin (in press) found that isolates that were presented in a larger font size than the rest of the list elicited a fronto-central positivity with the morphology of the Novelty P3, whose amplitude was correlated with subsequent recall. In line with this finding, Butterfield and Mangels (2003) found that a frontal positivity elicited by negative feedback about the participant's answer to a trivia question (which was followed by the correct answer) was correlated with whether or not they remembered the correct answer on a later surprise test.

Overall, the current theories on Novelty P3 function and the previously published data strongly diverge in regards to whether the Novelty P3 is functionally a direct link in the stimulus-response stream. Therefore, more research is needed to explore the relationship between Novelty P3, immediate behavioral responding, and subsequent memory.

The Pupil Dilation Response (PDR)

The variation of pupil size with cognitive processes has been studied for decades. Kahneman (1973) proposed pupil size as an index for task-related arousal, or "effort". In his (and others') studies pupil size was typically examined by comparing differences between tasks or by investigating tonic changes in pupil size across an entire task block rather than individual trials (e.g., Peavler, 1974). In contrast, the main focus of the present study is a phasic, temporary dilation of the pupil that peaks between 1s and 2s after stimulus onset (Friedman, Hakerem, Sutton, & Fleiss, 1973), the pupil dilation response (PDR).

The PDR as Part of the Orienting Reflex. The cognitive PDR elicited by non-noxious, novel stimuli is traditionally grouped together with a number of other physiological responses known under the umbrella term *orienting reflex* (Sokolov, 1963). While the other physiological responses are not directly investigated in the present study, it is useful to briefly discuss the phenomenon of the orienting reflex.

The orienting reflex was first reported by the Russian scientist Sokolov, who suggested that a collection of autonomic nervous system reactions to non-noxious stimulation reflect a "what is it" reaction in the organism (Kimmel, 1979; Sokolov, 1963). The most commonly studied component of the orienting reflex is the skin

conductance response (SCR): When a stimulus is encountered, skin conductance quickly increases and then returns to baseline. Upon subsequent presentations of the same stimulus, SCR amplitude habituates. When a dissimilar stimulus is encountered, a stronger SCR is elicited again ("response recovery"). Finally, a subsequent presentation of the "frequent" stimulus presented before will again elicit a strong SCR ("dishabituation") (e.g., Kimmel, 1979; Waters & Wright, 1979). The SCR is also sensitive to stimulus significance and intensity (Waters & Wright, 1979).

Other physiological responses classified under the orienting reflex concept include a deceleration in heart rate, an increase in blood pressure, a reduction in the power of the alpha frequency of the scalp-recorded EEG ("alpha blocking"), and the PDR (Barry, 2009). There has been a long debate of whether the P300 reflects a central nervous system analogue of the psychological process manifested in the orienting reflex (Donchin et al., 1984). The issue is complicated by the fact that in this literature the P300 and the Novelty P3 are rarely treated as distinct ERP components, conceptualizing the components together as the "P3 family", or as "subcomponents" of the LPC (for an example, see Rushby & Barry, 2007).

The original conceptualization of the orienting reflex as a collection of autonomic responses that all serve a common function is complicated by the heterogeneity of the sensitivity of each measure to experimental manipulation. In fact, there does not seem to be a single pair of autonomic responses that exhibits equivalent eliciting conditions. For example, only the SCR is sensitive to stimulus intensity (Barry, 2009). As another example, heart rate deceleration shows habituation, but does not show response recovery to stimulus change. However, although different autonomic components of the orienting

reflex might serve different functions, there appears to be some agreement that the orienting reflex reflects aspects of stimulus processing or response preparation that are integrated into the stimulus-response stream (e.g., Näätänen, 1978).

Eliciting Conditions of the PDR and their Similarities to the P300. The first study on the PDR was published by the same laboratory that originally discovered the P300. Thus, similarly to Sutton et al.'s (1965) study of the P300, Friedman and colleagues (1973) presented subjects with an auditory "click" that was followed by a second click at a certain probability. Before each trial, participants "guessed" whether a second click would occur. Both the P300 and the PDR elicited by the second click were inversely correlated with the probability of a second click to occur.

The PDR is also elicited by infrequent, task-relevant events in oddball paradigms (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Murphy, Robertson, Balsters, & O'Connell, 2011) and is equally enhanced when an outcome is better- or worse than expected (Preuschoff, 't Hart, & Einhauser, 2011) suggesting that the PDR, like the P300, is elicited by events that violate expectancies.

Further parallels between the two physiological responses are found in recognition memory tests, where previously studied ("old") items elicit both larger parietal ERPs (for a review, see Rugg & Curran, 2007) – in our view an instance of the P300 – and larger PDRs (Goldinger & Papesh, 2012; Heaver & Hutton, 2011; Otero, Weekes, & Hutton, 2011) than unstudied foils: The "P300 old/new effect" and the "pupil old/new effect" respectively. Finally, erroneous responses in a reaction time tasks elicit a P300 (Arbel & Donchin, 2009; Wessel, Danielmeier, & Ullsperger, 2011), as well as a PDR (Wessel et

al., 2011) and one recent study reported that both physiological responses are larger for perceived than unperceived errors (Wessel et al., 2011).

An additional, more indirect connection comes from each response's correlation to another ERP component that indexes violations of semantic expectations: The *N400* (Kutas & Hillyard, 1980). The absolute amplitude of the N400 has been shown to negatively correlate with both the PDR (Kuipers & Thierry, 2011) and the P300 (Arbel, Spencer, & Donchin, 2010). In summary, the PDR and the P300 for a larger part share their antecedent conditions, which has led to the idea that both physiological responses index the same psychological process (Nieuwenhuis et al., 2011). If this was the case their amplitudes should be correlated with each other. However, a recent study, applying a within-subject analysis, found no correlation between P300 and PDR (Murphy et al., 2011).

There are also important dissociations between the eliciting conditions of P300 and PDR. For example, in Stroop tasks, a larger PDR is elicited by incongruous, compared to congruous or neutral word-color combinations (Laeng, Ørbo, Holmlund, & Miozzo, 2011). This same effect is, however, not obtained for the P300 (Rosenfeld & Skogsberg, 2006). The prominent theory of Stroop interference attributes the increased reaction time to incongruous stimuli to response interference rather than stimulus processing (e.g., Ila & Polich, 1999), suggesting that the PDR may be more closely related to response-related processes elicited by unexpected stimuli than the P300.

The LC-NE Theory of the PDR and the P300. Although pupil diameter is not directly controlled by locus coeruleus (LC) activity, both receive afferent neural connections from the medulla. Furthermore, pupil diameter is correlated with single-unit

activity in the monkey LC, suggesting a close relationship between pupil dynamics and neural activity in the LC (cf. Nieuwenhuis, Aston-Jones, et al., 2005; Nieuwenhuis et al., 2011).

The well-supported "adaptive gain theory" (for a review, see Aston-Jones & Cohen, 2005) suggests that the LC controls behavior through an adjustment of its tonic and phasic activity levels. In its tonic mode, the LC exhibits high tonic ("baseline") activity and only small phasic responses to task-relevant stimuli. This mode corresponds to a state of "exploration", in which the individual seeks out new sources of potential reward within the environment rather than showing a narrow focus on attention on the task at hand. In contrast, the phasic mode, with low tonic, but strong phasic activity to task-relevant events, corresponds to a state of "exploitation", in which the individual exhibits strong task focus, thus pursuing a known source of reward.

There is evidence that pupil size patterns reflect the tonic vs. the phasic LC mode: In time periods where baseline pupil diameter is relatively small, but phasic PDRs are large (phasic mode), the participant is strongly engaged in the task at hand and shows good performance. In contrast, when baseline diameter is relatively large, but the PDR is rather small (tonic mode), participants are disengaged from the task at hand and show rather poor performance (Gilzenrat et al., 2010; Jepma et al., 2011). Through elegant supplemental analyses, Gilzenrat and colleagues (2010) ruled out the possibility that this difference in the phasic PDR is due to a ceiling effect in pupil size due to the increased tonic activity. Combined with evidence for a role of the LC in adapting behavioral focus (see Aston-Jones & Cohen, 2005), this suggests that the phasic PDR may be related to aspects of immediate responding.

The same group of authors (Nieuwenhuis et al., 2011) have proposed that, like the PDR, the P300 reflects phasic activity in the LC, and both are therefore closely related to behavioral performance on the same trial. In support of this idea, Murphy et al. (2011) found that time periods with smaller baseline pupil diameters were associated with both larger PDRs and larger P300s. Importantly, however, on an individual trial basis, there was no correlation between the amplitudes of the PDR and the P300, casting some doubt on the idea that PDR and P300 reflect the same psychological function.

The PDR and Reaction Time. According to the same logic outlined for the P300, if the phasic PDR indexes aspects of behavioral responding, its latency should be correlated more closely with reaction time than stimulus onset. Indeed, one prior study found that, across participants, PDR peak latency and reaction time were positively correlated (Nuthmann & Van Der Meer, 2005). However, stronger evidence for an association would be obtained in a within subject, individual-trial analysis, and thus far no studies have investigated this.

Although there are not many prior studies on the correlation between PDR *latency* and reaction time, there are several studies that investigated PDR *amplitude* and reaction time. For example, the PDR is enhanced for incongruent trials in the Stroop task, a trial type that also leads to increases in reaction time (e.g., Laeng et al., 2011; Siegle, Steinhauer & Thase, 2004). The source of the Stroop interference is often attributed to a conflict in response preparation (e.g., Ila & Polich, 1999), as opposed to stimulus processing, providing some indirect support for the idea that the PDR indexes response-related processes.

Stronger evidence for an association between PDR amplitude and reaction time comes from Gilzenrat et al. (2005, exp. 1), who found that in an oddball task, the phasic PDR was larger in trial types in which reaction times were short and response accuracy was high (note, however, that the authors did not study this relationship directly, but by relating both enhanced performance and larger PDR to smaller baseline pupil diameters). In a similar experiment, Murphy et al. (2011) found that the amplitude of the PDR was larger in situations in which participants performed relatively *poorly*. However, large PDRs were followed by an improvement in performance, indicating that behavioral adjustments in response to novel events are reflected in the PDR.

In summary, there is some evidence for a relationship between the PDR and behavioral performance in the present- or immediately following trials. This idea is consistent with a prevalent theory of PDR function, the adaptive gain theory. Furthermore, even Kahneman's original idea attributing pupil dilation to mental "effort" is consistent with the association of pupil dilation with immediate responding.

The PDR and Memory Encoding. A classic discovery relating pupil size to working memory load comes from the digit-span task (Kahneman & Beatty, 1966). Thus, the pupil continuously dilates while participants encounter a number of successively presented digits ("loading" phase), which are to be retained for subsequent immediate serial recall. As the participant recalls the stimuli ("unloading" phase), the pupil diameter continuously decreases. This phenomenon is traditionally interpreted in terms of mental "effort" – the more numbers are to be retained in working memory, the higher is the participant's effort to retain these stimuli, and hence the pupil diameter constantly increases until the working memory capacity is reached. Then, as the numbers are being

recalled, the working memory load and thus effort decreases, leading to the decrease in pupil size (for a review, see Kahneman, 1973).

The relationship between PDR and episodic memory has been most frequently studied by recording the PDR during a recognition test. These experiments have consistently reported a "pupil old/new effect" – larger PDRs for previously studied stimuli compared to non-studied foils (e.g., Otero et al., 2011).

There are few prior studies on the relationship between the PDR and episodic memory using the subsequent memory paradigm, and the results are inconsistent (for a review, see Goldinger & Papesh, 2012). Papesh and Goldinger (2011) reported that auditory study words that were subsequently successfully recognized with high confidence had elicited larger PDRs at study than words recognized with low confidence, or forgotten words. In contrast, Kafkas and Montaldi (2011) reported that incidentally encoded pictures that were later judged as more familiar elicited *smaller* PDRs at study than those that were judged as less familiar. Similarly, Nabler and colleagues (2013) found that larger *constrictions* to the onset of pictures predicted subsequent recognition. It is important to note, however, that since photographs show different luminance characteristics than words, this constriction effect might be specific to pictorial stimuli. Overall, more research is necessary to investigate subsequent memory effects in the PDR.

In summary, several prior studies have linked pupillary responses to behavioral performance on the same- or the following trials. These findings are in line with the idea that the PDR might be directly integrated into the stimulus-response stream. However, prior reports on subsequent memory effects are rare and inconsistent, so the role of the PDR in memory encoding is to date unclear.

The Present Study

A review of the literature suggests that there is much overlap in the eliciting conditions of the P300, Novelty P3 and PDR. However, although it appears that P300 may be related more closely to memory encoding than behavioral responding, and that the reverse is true for the PDR, the extent to which each response indexes immediate responding and/or memory encoding remains controversial. The goal of the present study is to investigate this issue using a modified Novelty P3 oddball paradigm. We conducted a thorough study of the correlation of each response with behavioral measures, and an additional regression analysis on individual trials aimed to determine the extent to which the physiological variables remain predictive of the behavioral measures *when variance of other physiological responses has been accounted for*. Furthermore, by recording multiple physiological indices of novelty, more general effects of "attention" as a third variable influencing both physiological responses as well as memory encoding can be ruled out (as these would be expected to affect *all* responses).

Our design was similar to the typical Novelty P3 oddball paradigm, with the major modification that the pictures inserted into the sequence (i.e., the "novels") were classifiable along the same dimension as the other stimuli. This design was chosen because it allowed us to study the relationship of the elicited physiological responses to immediate responding. Each Novelty oddball task was followed by recall- and recognition tests. We thoroughly studied the behavioral data obtained from each phase of the experiment, especially the extent to which behavior varied between stimulus types, in order to aid an interpretation of the variance of each physiological response with stimulus type. We studied the relationship between each physiological response with (1) reaction

time on the same trial, (2) subsequent recall, (3) subsequent recognition speed and (4) reaction time on the next trial. To do so, we performed both median split analyses on reaction times and an analysis of individual trials. Finally, a regression analysis attempted to determine the correlation with each measure and performance when the other measures had been accounted for.

Hypotheses. While we also thoroughly studied patterns in the behavioral data in our paradigm, our main hypotheses addressed the variance of the physiological responses with experimental manipulations and behavior. The hypotheses concerning the P300 and the PDR were directed, while the Novelty P3 hypotheses were more exploratory due to a scarcity of relevant prior findings.

Based on a wealth of prior data, we hypothesized that a P300 and a PDR will be elicited by the infrequent task-relevant category. Furthermore, we expected that infrequent presentations of pictures ("novels") among the verbal stimuli would elicit a Novelty P3, a P300, and a PDR.

Secondly, we expected that the P300 would exhibit a subsequent memory effect, with larger amplitudes for later recalled, compared to not recalled, infrequent stimuli. The hypothesis about the correlation between P300 amplitude and subsequent recognition speed was more exploratory – prior studies have not found a P300 subsequent memory experiment when recognition memory was tested (Fabiani et al., 1990). However, since we focused on *reaction time* rather than accuracy during recognition, we expected that a P300 subsequent memory effect might be observed.

Since we hypothesized that the PDR indexes processes related to immediate responding, we predicted that its amplitude and latency would correlate with reaction

time on the same trial. Furthermore, based on the literature reviewed above, we also predicted that there would be a correlation between PDR measures and reaction time to the immediately following trial.

Finally, we hypothesized that while the P300, Novelty P3 and PDR are elicited by the same stimuli, they represent separate psychological processes, and therefore the measures would not strongly correlate with each other on an individual trial basis. We also expected that all physiological indices would be correlated with both reaction time and memory, but a regression analysis attempted to reveal whether each physiological response predicts the respective outcome (reaction time or memory) when the other physiological responses have been accounted for.

Methods

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

Participants

In exchange for partial course credit, twenty-nine healthy undergraduate college students from the University of South Florida Psychology Department's participant pool took part in this experiment, which was part of a larger, two-session study[1]. One participant was excluded because their memory performance (recall and recognition) deviated from the sample by more than 2 standard deviations. Furthermore, 3 participants were excluded due to technical difficulties during the participant run. The final sample included 25 participants, aged 18-49 years (M=23.44). Six participants were male and 4 were left-handed. All participants were native speakers of English with normal or corrected-to-normal vision.

Stimuli

The stimuli presented in the encoding phase of each experimental block were classifiable according to one of three rules: (1) edible vs. inedible, (2) living vs. non-living and (3) smaller vs. larger than a shoebox. Lists of nouns for each of the 6 categories were created by drawing words from Francis and Kucera's (1982) database. In

[1] The other session was scheduled one week apart and involved an experiment studying ERPs and pupil responses elicited during motor-preparation. The main purpose of using the same participants in two experiments was to reduce the amount of preparation time required for the calibration of the eye tracker.

an initial selection step, six college students were asked to apply the classification rules (e.g., edible vs. inedible) to the respective set of words. Words for which these participants indicated that they couldn't easily be classified or that their category membership was ambiguous were excluded.

We next we selected line drawings of a subset of the words. Thus, for as many words as possible, we obtained Clip Art pictures that depicted items referred to by the words and converted each image to black color. We pre-selected about 40 images of each category that, according to our judgment, depicted most clearly the respective noun. These pre-selected pictures were then shown in random order to a sample of 20 participants (none of which later participated in the main experiment) in a paper-based questionnaire.[2] For each image the participants were asked to (1) name the object depicted and (2) rate the difficulty of naming the image on a scale of 1 ("not difficult") to 7 ("very difficult"). We then selected the 20 images of each category for which the provided names showed the largest overlap between participants and that were also rated as easy to name. Note that very similar labels like "bread" vs. "loaf of bread" or "glasses" vs. "eye glasses" were counted as the same label. For the final set of the 20 pictures of each category, at least 18 of the 20 participants had provided the same label, and the average naming difficulty rating (on a scale of 1 to 7) was less than 1.8 (M=1.17). The words corresponding to the selected images in each category were then removed from the word lists.

From the remaining words in each category, 136 nouns including between 3 and 9 letters were selected. Care was taken to match word frequency (lemma occurrences per

[2] The participants in this pilot procedure were between 16 and 60 years old and all reported English to be their first language. 11 were female and 9 were male.

million from Francis and Kucera, 1982) and length (number of letters) of the words in the edible (frequency: 1-78, M=11.49; length: M=5.46) to the inedible (frequency: 1-48, M=12.35; length: M=5.49) category, the living (frequency: 1-100, M=22.45; length: M=5.6) to the non-living (frequency: 1-77, M=20.46; length: M=5.62) category, and the smaller (frequency: 1-99, M=16.36; length: M=5.59) to the larger than a shoebox (frequency: 1-94, M=16.98; length: M=5.68) category.

The final lists of stimuli contained 136 nouns and 20 images for each of the six categories. The participants were highly accurate in categorizing the stimuli (M=.94, see behavioral results for details), suggesting that the selected stimuli were appropriate for the present task and sample.

All stimuli were presented in black font on a light grey background (RGB values 125, 125, 125). The words were displayed in font size 40 in Arial font, therefore spanning between 2.8 and 8.6 degrees of the visual angle. The largest dimension (either width or height) of the pictures was 95mm, and therefore in their largest dimension pictures spanned 8.6 degrees of the visual angle.

Task and Procedure

Every experimental session took place at 9am to avoid arousal changes due to time of the day as a nuisance variable. All participants first gave informed consent, after which the electrode net was applied. The participant then took a seat at a distance of about 60cm in front of a computer screen, after which the experimenter calibrated the eye tracker. The preparation time for the EEG- recording and eye tracking was up to 30min.

The experiment contained a practice block and six experimental blocks, each consisting of an encoding phase, immediate free recall, a distraction task and a

recognition phase. Figure 1 illustrates the structure of an experimental block. After the final block, participants were debriefed and given information about the purpose of the experiment. The duration of the experiment did not exceed 2 ½ hours.

Encoding. At encoding, participants completed an oddball task that involved one of three types of semantic judgments, each of which was (along with the corresponding stimulus set) randomly assigned to two successive experimental blocks. That is, blocks 1 and 2 shared the same task, as did blocks 3 and 4, as well as blocks 5 and 6. Therefore, we will refer to blocks 1, 3, and 5 as "sub-blocks 1" and to blocks 2, 4 and 6 as "sub-blocks 2". The three semantic judgment tasks were:

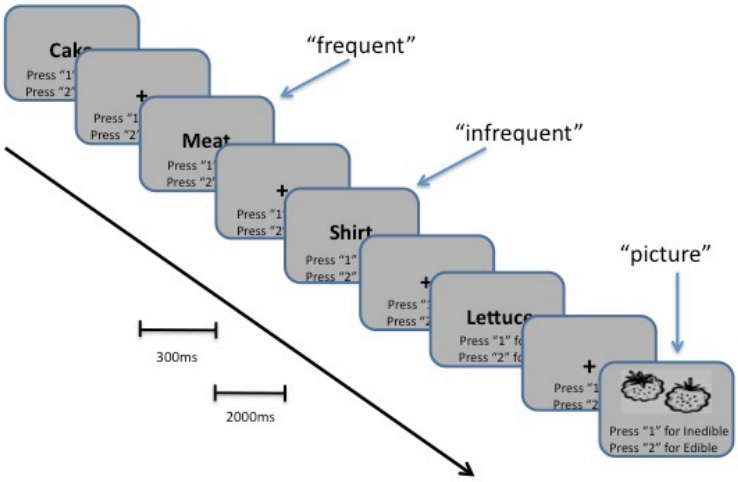
(1) Living vs. nonliving (e.g., lion vs. pencil)

(2) Smaller vs. larger than a shoebox (e.g., ant vs. ship)

(3) Edible vs. inedible (e.g., pizza vs. table)

The order of the tasks was randomized. Participants pressed one of two buttons with their left or right hand to classify each stimulus. Response hands were assigned randomly for each block, and the assignment was displayed on the bottom of the screen throughout the task (figure 1). While participants were informed that memory tests would follow, they were instructed to focus their full attention on the semantic judgment task and to respond as quickly and accurately as possible. Since the stimuli were presented in a relatively quick sequence and each stimulus required a response, we assumed that participants were not using elaborative memorization strategies in parallel to performing the encoding task.

1. Encoding (duration: 2 min 48s)

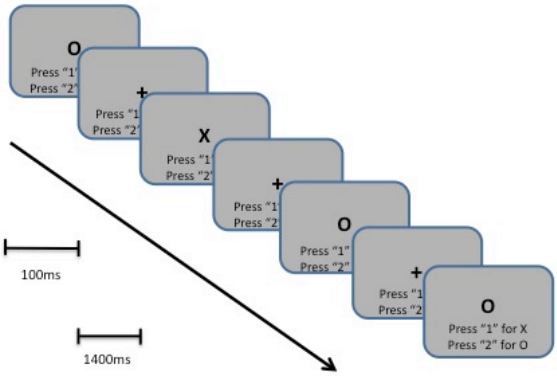


One of three semantic judgment tasks:

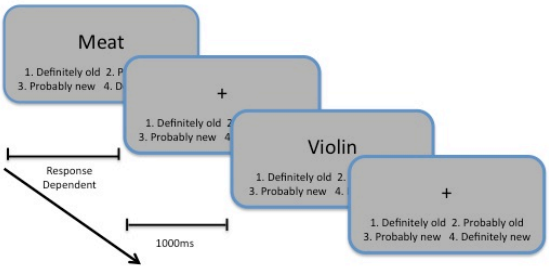
1. Edible vs. inedible
2. Living vs. non-living
3. Smaller vs. larger than a shoebox

2. Recall (duration: 3 min)

3. Distraction (duration: about 3 min 30s)



4. Recognition



Presented are:

- all "old" infrequents and pictures
- 10 "old" frequents
- an equal number of "new" infrequents, pictures and frequents

5. Performance Feedback

Figure 1. Structure of one experimental block. Shown is a second sub-block, which included pictures in the encoding sequence. See text for details.

Each encoding phase contained a sequence of 73 stimuli, which were presented for a duration of 300 ms each, followed by a 2000 ms-long fixation cross. In blocks 1, 3 and 5 (henceforth referred to as "sub-blocks 1"), 63 of these stimuli were of the frequent-, and 10 were of the infrequent category; for example, 63 words (86%) may have been "edible" while 10 words (14%) were "inedible". Blocks 2, 4, and 6 (henceforth referred to as "sub-blocks 2") contained 53 words of the frequent category (72%), 10 words of the infrequent category (14%) and 10 line drawings of items in the frequent category (14%). Which category was infrequent within each block was randomized across participants. Pictures were to be classified according to the same task as the words (and hence, the pictures always required the frequent response – this was, however, not explicitly stated to the participants). The first three and the last three stimuli were always words of the frequent category, and these words were not further analyzed or presented in the recognition phase. Each stimulus was drawn at random from the respective lists, with the restriction that no two infrequents and no two pictures could be presented successively. Stimuli not used in the encoding phase served as foils in the recognition task (see below).

The purpose of including the line drawings in every other task block was to elicit a Novelty P3. It is worth noting, however, that these "novels" are very different from the ones used by Courchesne et al. (1975) in that they were images of familiar items, and in that they required a response. The rationale for using classifiable stimuli was that we intended to determine the relationship between each physiological response and reaction time on the same trial. It is worth noting that in prior studies, pictures of familiar objects (Cycowicz & Friedman, 2007) and novels that were task-relevant (Cycowicz &

Friedman, 1999, 2004) have elicited a Novelty P3. Combined with pilot data for the present design, this suggested that our paradigm was suitable to elicit a Novelty P3.

Recall. Immediately after each encoding phase (i.e., each set of 73 stimuli), participants were asked to write down, in any order, every word or picture label they remembered from the preceding task. The amount of time allowed for the recall phase was 3 minutes. At the end of the recall phase, an experimenter entered the room to take away the recall sheet.

Distraction Phase. The recall phase was followed by a distraction task lasting about 3.5 minutes. The distraction phase consisted of a simple oddball task in which participants had to press one of two buttons according to whether each of a sequence of 140 stimuli was an "X" or an "O". Each stimulus was presented for 100ms, followed by a 1400ms long fixation cross. The probability of the infrequent stimulus on any given trial was .2, and the assignment of X or O as the infrequent was randomized. Data from the distraction phase are not reported in the present paper.

Recognition. Immediately after the distraction task the recognition phase began. In each trial, participants judged their recognition memory for the respective stimulus, by pressing one of four buttons, on a scale of 1 ("definitely old") to 2 ("probably old") to 3 ("probably new") to 4 ("definitely new"). The response assignment was displayed at the bottom of the screen the entire time. All infrequents (n=10) and all pictures (n=10; only in blocks 2, 4, and 6), as well as a random sample of 10 frequents from the encoding phase were tested, along with an equal number of unstudied stimuli drawn from the same pool. Presentation order was random. The stimulus stayed on the screen until the

response was given and between two recognition trials a fixation cross was presented for 1 sec (figure 1).

Performance Feedback. After the conclusion of each block participants were given feedback about their performance. That is, their classification accuracy and average response time at encoding, as well as their recognition rate (i.e., the proportion of correct responses at test; this evaluation ignored the confidence judgment), was displayed on the screen. After receiving this feedback, participants were allowed to take a break for as long as they wished.

Practice. Before the first experimental block participants completed a practice, which was shorter than the experimental blocks but followed the same structure. In the practice encoding phase, participants judged 10 names according to whether they were male or female names, the recall phase lasted only 45 seconds, the distraction task included only 10 X/O oddball trials, and the recognition phase included a random sequence of only 5 "old" and 5 "new" names. The instructions were the same as for the remainder of the experiment, and the purpose of the practice was to insure that participants were familiar with the task structure and understood all instructions.

EEG Recording and ERP Analysis

The EEG was recorded with a 128 electrode EGI system, with the central electrode (Cz) as the on-line reference site, and digitized at a sampling rate of 250 Hz. Using Netstation software, we off-line band-pass filtered the EEG with cutoff frequencies of 0.3 and 20 Hz, and replaced bad channels by a mathematical interpolation procedure. Then, the EEG was sliced into segments of 300 ms before- to 1400 ms after stimulus onset. The extracted stimulus categories included frequents presented in the first sub-block,

frequents in the second sub-block, infrequents in the first sub-block, infrequents in the second sub-block and pictures. Note that in order to avoid as best as possible influences of sequential effects on the physiological measures (K. C. Squires et al., 1976), frequent stimuli were only included if they were preceded by frequents. Eye blinks were removed by an independent component analysis (ICA) technique provided by Joe Dien's toolbox (Dien, 2010a) in Matlab. The eye blink-corrected trials were then re-referenced to linked mastoids and automatically and visually screened for artifacts. Any trials including artifacts were excluded from further analysis. Finally, subject ERPs were calculated for each stimulus category.

The subject ERP averages for all stimulus types were submitted to a spatial PCA (Spencer et al., 1999) to identify ERP components and to obtain "factor scores" as amplitude measures. PCA factors that exhibited the morphologies typical for the Novelty P3 or the P300 were selected for further analysis. Often, spatial PCA is followed by a temporal PCA step to extract temporal patterns in the data. However, this method provides one temporal factor with a fixed latency to represent each ERP component of interest and is therefore inappropriate to measure latency differences within an individual component (e.g., to compare latencies between trial types or to quantify latency for individual trials). Rather than using temporal PCA, we therefore quantified ERP component amplitude as the maximum factor score in the baseline corrected "virtual ERPs" (spatial factor scores over time, Spencer et al., 1999) within a specified time window, as reported in the results section. Component latency was defined as the time point (in ms) of the maximum factor score. Such measures have previously been used successfully (Brumback, Arbel, Donchin, & Goldman, 2012).

Quantification of ERP components in individual trials was equivalent to quantification for the averages: The EEG was "filtered" through the respective spatial factor coefficients to obtain "virtual ERPs", and then amplitude and latency were extracted by using the peak picking procedure on the individual trial's virtual ERP.

Pupil Size Recording and PDR Analysis

Pupil diameter was recorded from both eyes at a sampling rate of 60 Hz, using SmartEye Pro 5.8 software and two cameras installed below the participant screen. For all off-line analysis, we used self-written Matlab code. The pupil diameter recording was first sliced into segments from 1000 ms before- to 3000 ms after stimulus onset, using the same trial categories as for the ERP analysis (although we immediately collapsed across sub-blocks for the pupil diameter analysis). Data points for which the pupil diameter measurement was below 2mm or above 10mm, or for which the pupil diameter deviated from the average of the previous 3 data points by more than 0.5mm were marked as bad because they were physiologically implausible. Note that this procedure also marks eye blinks as bad data points. Bad data points were then replaced by a linear interpolation procedure using the values of the two "good" data points that immediately surrounded bad data points. Trials were excluded from further analysis if more than 25% of the data points, or 15 sequential data points were marked as bad, because in these cases the interpolation procedure was not expected to deliver reliable results.

Next, a single pupil diameter recording was obtained for each time point by averaging across the measurements from both eyes. A 3-point moving average filter was then applied to the data. Subject averages were computed for each trial type, including 500ms

before to 2000 after stimulus onset. Finally, each subject average was baseline corrected using the 500ms before stimulus onset.

The pupil dilation response (PDR) analysis – both for the averages and for the single trials – involved extracting the maximum change in diameter, compared to the pre-stimulus baseline, within a time window of 1000 to 1500ms after stimulus onset. Latency was defined as the time point (in ms) of the maximum. In addition, we analyzed the mean amplitude in the time window where pupil size returned to baseline (1500-2000ms after stimulus onset), as well as the absolute baseline pupil diameter.

Results

We begin by reporting the behavioral data, which will aid in a characterization of the task demands imposed by our design, as well as constrain subsequent interpretations of the functional significance of the elicited physiological responses. Furthermore, analysis of the behavioral data also provides suggestions for which covariates to include in our regression models. Second, we will report an analysis of the ERP data using spatial PCA. We will compare component amplitudes and latencies between stimulus types (frequent vs. infrequent vs. picture), report a subsequent memory analysis using the typical approach, as well as an median-split based analysis of component sensitivity to reaction times on the same trial, during the recognition test, and to the successive trial. Third, we will report a pupil size analysis, performing the same statistical comparisons as for the ERPs. In the final section we report the regression analysis, which brings together data from all previous sections.

Behavioral Data

Analysis of the behavioral data included an analysis of (1) reaction time and accuracy at encoding, and (2) recall rates, recognition accuracy, and reaction times during the recognition test. These analyses will reveal differences in stimulus processing/response preparation time (as indexed by reaction time at study) as well as task difficulty (as indexed by a combination of reaction time and accuracy at study) between encoding tasks, task blocks and stimulus types. Furthermore, it will reveal differences in strength or efficiency in encoding and retrieval as it varies with the different conditions. Outlining

these differences will subsequently aid in an interpretation of the variance in the physiological measures with the different conditions (especially between frequents, infrequents and pictures).

Behavioral Data: Reaction Times and Accuracy at Encoding. Our behavioral measures for the encoding phase included median reaction time and the proportion of accurate responses. Overall, participants were highly accurate in performing the encoding tasks (accuracy M=.94), suggesting that our design was suitable and the selection of our stimuli was appropriate for our sample. For the analysis of the reaction times it was useful to first examine the *shape* of the reaction time distribution.

Therefore, we first constructed a vincentized group reaction time curve across all trial types from the encoding phase (the vincentization procedure "summarizes" the shape of a distribution across participants; it calculates quantiles for the group distibution by averaging over the participants' individual quantiles; see Ratcliff, 1979). The vincentized probability density function (figure 2A, top panels) was positively skewed and therefore showed the typical shape of reaction time curves in two-choice reaction time tasks (e.g., Ratcliff, 1979).

For the purpose of a regression analysis it was useful to transform the data to a distribution that more closely approximates normality, since with normally distributed data it is more likely that the residuals are also normally distributed – a key assumption of regression. To this end, we first excluded outliers for each participant by eliminating reaction times that deviated by at least 3 standard deviations from their mean reaction time. Afterwards, we log-transformed each data point; such a transformation tends to

normalize reaction time data (Ratcliff, 1993). Indeed, as shown in figure 2A (bottom panels), the resulting reaction time distribution approximated normality more closely.

Due to the skew, we compared reaction times between conditions and stimulus types using the median (rather than the mean) as the measure of central tendency. Statistically, we analyzed differences with two-sided paired samples t-tests or repeated measures ANOVAs.

Performance in the Three Different Encoding Tasks. When we designed the three different semantic judgment tasks ("edible": edible vs. inedible, "size": larger vs. smaller than a shoebox, "living": living vs. non-living) and selected the corresponding stimuli in each category, we assumed that the three tasks would be about equivalent in their task demands. To test whether this assumption was correct we tested for differences in the median reaction times between tasks (figure 2B). In the contrary to our expectation, there was an overall difference in reaction time [$F(2,48)=4.44$, $p=.02$] between the "size" task (M=657.52ms), the "edible" task (M=618.68ms) and the "living" task (M=638.8ms). Post hoc t-tests suggested that reaction times in the "size" task were significantly longer than for the "edible" task [$t(24)=2.9$, $p<.01$], while no other differences were significant ["edible" vs. "living": $t(24)=-1.94$, $p=.07$; "size" vs. "living": $t(24)=1.26$, $p=.22$].

When response accuracy was collapsed across stimulus types (frequents, infrequents and pictures), there were no differences between the three task types ($p>.9$). However, differences became apparent when only trials of the infrequent category were compared [$F(2,48)=3.32$, $p<.05$]. Most likely, ceiling performance for frequents (M=.96) and for pictures (M=.98) prevented the detection of differences when accuracy was collapsed across stimulus types. Thus, infrequents in the "size" task were associated with lower

accuracy (M=.72) than infrequents in the "edible" task (M=.82) [$t(24)=2.26$, $p=.03$], but neither the difference between infrequents in the "size" and the "living" category (M=.79) [$t(24)=1.98$, $p=.06$], nor the difference between "edible" and "living" [$t(24)=.73$, *ns*] was significant. Taken together, the reaction time and accuracy data suggest that the "size" task was more difficult than the "edible" task. One implication of this difference is that task type should be included as a covariate in the regression analyses of reaction time at encoding.

Performance Throughout the Experiment. It is possible that performance changed across the course of the experiment due to fatigue or learning effects. To investigate this, we tested for reaction time differences between blocks 1 and 2 (M=647.24ms), blocks 3 and 4 (M=636.76ms), and blocks 5 and 6 (M=631ms; figure 2C). While there was a slight tendency for response times to decrease across the course of the experiment, the difference was not significant [$F(2,48)=.69$, *ns*.]. Likewise, there were no significant differences in response accuracy; not even when only accuracy for infrequent stimuli was analyzed [$F(2,48)=1.61$, *ns*].

Performance in the First and Second Sub-blocks. Next, we investigated whether performance differed between those task blocks that included only words of the frequent- and the infrequent category (sub-blocks 1), and blocks that also included pictures (sub-blocks 2). We expected that *if* there were any differences, sub-block 2 would likely show longer reaction times because the presence of the pictures might increase task difficulty. In contrary to this idea (figure 2D), reaction times were significantly longer for the first (M=652.04ms) than for the second (M=627.82ms) task block [$t(24)=3.27$, $p<.01$]. This difference cannot be solely driven by shorter reaction times to pictures, as pictures were

associated with numerically longer reaction times than frequents (see section on trial types below). Since two sub-blocks always shared the same semantic judgment task ("edible", "size" or "living"), the reaction time difference could reflect learning effects.

Response accuracy did not differ between the first and the second sub-block, neither when the analysis was collapsed over stimulus types [$t(24)=.26$, *ns*], nor when it was conducted separately for frequents (M=.94 for the first and M=.95 for the second sub-block) [$t(24)=.58$, *ns*] and infrequents (M=.79 for the first and M=.74 for the second sub-block) [$t(24)=1.66$, *ns*].

Performance for Frequents, Infrequents and Pictures. The analyses reported thus far tested for unintended effects of our study design on reaction times, which would need to be accounted for in further analyses. The manipulation of including frequents, infrequents and picture trials was, however, expected and intended to lead to differences in response characteristics (figure 2E). Indeed, median response times differed between trial types [$F(2,48)=45.63$, $p<.01$], such that participants were slower in responding to the infrequent category (M=731.26ms) than to frequents (M=622.52ms) [$t(24)=9.27$, $p<.01$] and pictures (M=637.48ms) [$t(24)=6.24$, $p<.01$], while pictures and frequents did not differ from each other [$t(24)=1.55$, *ns*].

Error rates also differed between trial types [$F(2,48)=85.16$, $p<.01$]: participants were more likely to make errors on infrequent trials (percent accurate M=.78) than for frequents (M=.96) [$t(24)=9.65$, $p<.01$] and pictures (M=.98) [$t(24)=10.73$, $p<.01$], with no difference between frequents and pictures [$t(24)=1.53$, *ns*]. The increased response times and error rates for infrequent trials could reflect a cost of response switching, as only infrequents required the infrequent response.

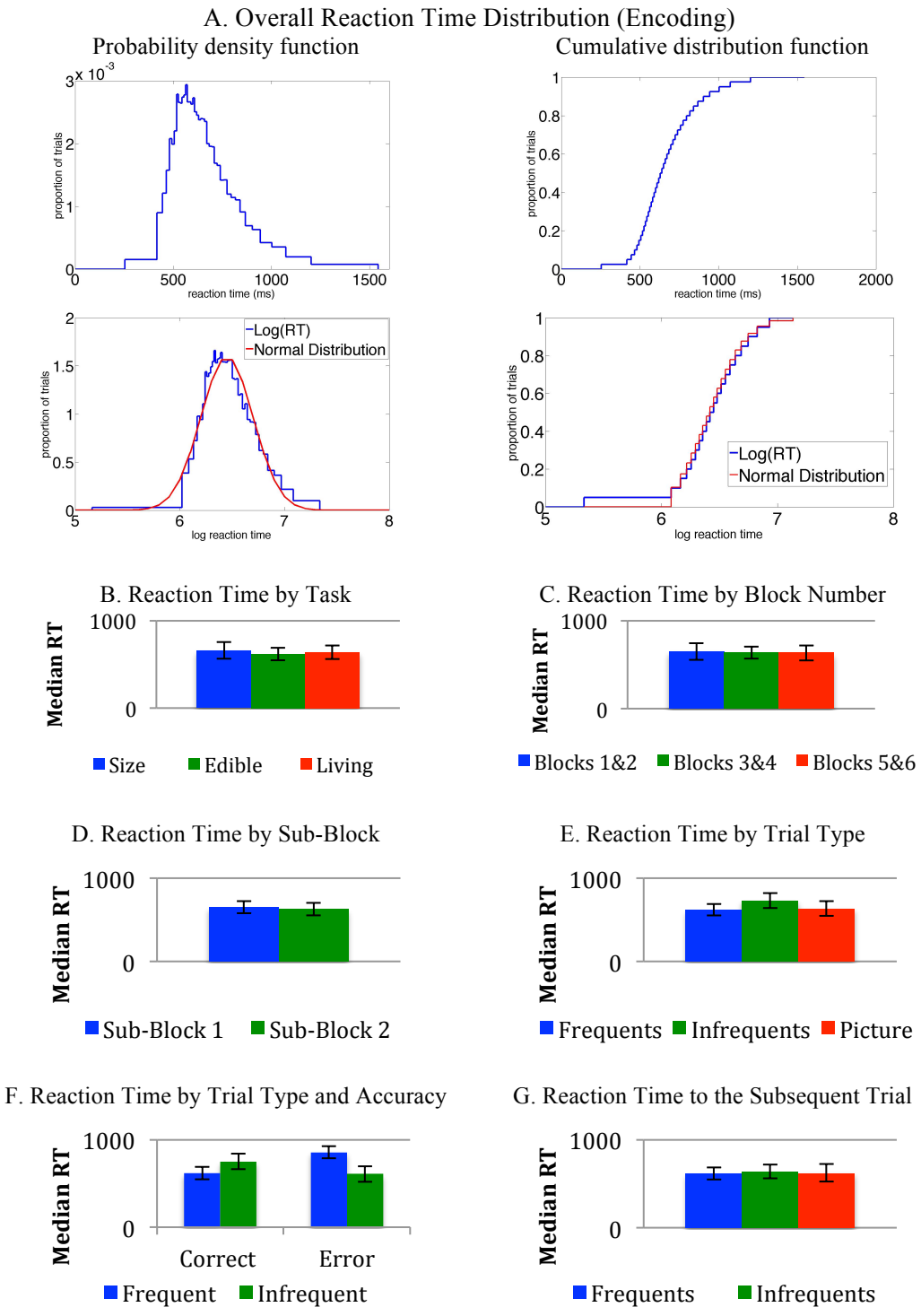


Figure 2. Reaction time data from the encoding phase. A. Overall vincentized, group-level reaction time (RT) distribution. Top panel: probability density function (pdf) and cumulative distribution function (cdf) of the raw RTs. Bottom panel: Log-transformed distribution, with a normal distribution overlaid. B-G: Median RTs and cdf's by task type (B), block number (C), sub-block (D), trial type (E), trial type and accuracy (F) and by previous trial type (G).

Error vs. Correct Trials. Much prior literature has suggested that response times can vary in systematic ways with response accuracy (Rabbitt, 1969). Since many participants committed no errors for picture trials, our analysis of this issue focused on trials of the frequent and the infrequent category (figure 2E). Thus, we submitted the median reaction times to a trial type (frequent vs. infrequent) by accuracy (correct vs. incorrect) ANOVA. While none of the main effects were significant ($p>.09$), we found a significant interaction [$F(1,23)=58.58$, $p<.01$]. As is clearly visible in figure 2E, for frequents, error trials (M=859.04ms) were slower than correct trials (M=618.9ms) [$t(24)=5.25$, $p<.01$], while for infrequents, errors (M=610.4ms) were faster than correct responses (M=754.06ms) [$t(24)=7.66$, $p<.01$]. This pattern suggests that for frequents, errors might be due to a "true" miscategorization of the respective word, with the longer reaction time indexing a "hesitation" to respond. In contrast, error responses to infrequents are most likely due to a premature execution of the frequent response before the stimulus has been fully evaluated.

Reaction Time to the Following Trial. A final behavioral measure during the encoding phase that was relevant to our hypotheses was the performance on the trial that *followed* frequents, infrequents and pictures. Only trials that were followed by frequent trials were included in this analysis. Based on the results by Notebaert et al. (2009), we expected that responses will be slowed down after the presentation of an infrequent event, which in our design might include infrequents and pictures. Indeed, infrequents (M=642.4ms) and pictures (M=625.94ms) tended to be followed by slower responses than frequents (M=619.14ms). However, this difference was not significant

[$F(2,48)=2.56$, $p=.09$]. Responses following frequents (M=.96), infrequents (M=.96) or pictures (M=.97) also did not significantly differ in response accuracy [$F(2,48)=.8$, *ns*].

Reaction Time and Accuracy at Encoding: Summary. The reaction times showed the typical, positively skewed, distribution. However, after excluding outliers and log transforming the reaction times, the distribution approximated normality. Our analyses of the median reaction times and response accuracies suggest that our experimental design had several unintended effects: Task performance differed between the semantic judgment tasks, and between blocks that did- and blocks that did not include pictures. These findings suggest that in the following analysis (in particular the regression analysis), task type and sub-block should be included as predictors of reaction time.

We also found that reaction times were slower and accuracy was lower for infrequents compared to both frequents and pictures. This difference is most likely due to the cost of response switching: Infrequents required a switch to the infrequent response, while neither frequents nor pictures did. Our design, however, confounds semantic deviance with a switch in the correct response, so an alternative possibility is that processing the semantic content of the stimuli of the frequent category (including the pictures) benefitted from conceptual priming effects. We will return to this point in the discussion.

Finally, errors were associated with slower responses than correct responses for frequents, while the reverse was true for infrequents. Since this interaction may complicate further analysis, trials with erroneous responses were subsequently excluded.

Behavioral Data: Memory Performance. The behavioral measures of memory performance included the proportion of stimuli recalled, recognition accuracy, as well as

reaction time during the recognition test. Note that the first three and the last three stimuli presented in the encoding phase were "buffer" stimuli (which were always of the frequent category), intended to absorb primacy- and recency effects, and were therefore not included in the analysis. Furthermore, it is worth noting that we will keep referring to the "frequent" and the "infrequent" categories, although during the recognition test all categories were equally frequent (i.e., only a subset of the frequents were presented during the recognition test). For instance, we will use the term "frequent" to refer to the category that was presented as the frequent category at *encoding*.

Recall Rates. Overall, participants recalled about 20 percent of the stimuli presented at encoding (M=.19). Recall rates did not differ between encoding tasks ("size": M=.2; "edible": M=.19; "living": M=.19) [$F(2,48)=.29$, *ns*], and did not decrease between blocks 1 and 2 (M=.2), blocks 3 and 4 (M=.19) and blocks 5 and 6 (M=.19) [$F(2,48)=.24$, *ns*]. However, there was a statistical trend for lower recall rates in the second sub-blocks (M=.18), compared to the first sub-blocks (M=.2) [$t(24)=2.06$, $p=.05$], suggesting a tendency for proactive interference for task blocks including stimuli of the same category.

As shown in figure 3A, there were differences in recall rates between stimulus types [$F(2,48)=18.78$, $p<.01$], indicating that frequents (M=.16) were recalled with a lower probability than infrequents (M=.27) [$t(24)=4.65$, $p<.01$] and pictures (M=.32) [$t(24)=6.83$, $p<.01$]. The difference between infrequents and pictures was not significant [$t(24)=.17$, *ns*].

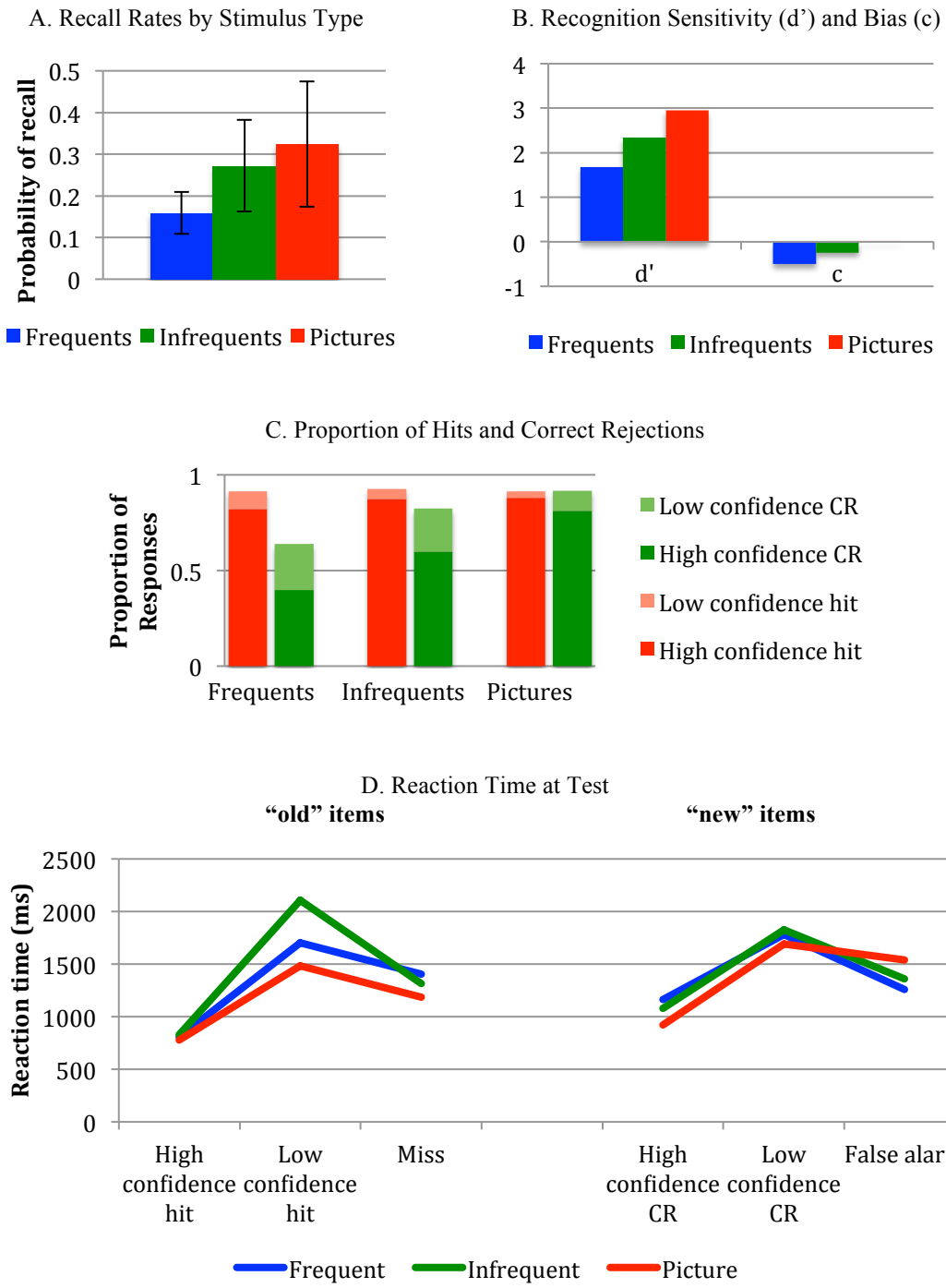


Figure 3. Memory performance. A: Recall rates for each trial type. B: Recognition accuracy (sensitivity, measured by d') and the bias measure (c) for each trial type. C: Proportion of old items that were given a correct response ("hit") and proportion of new items that were given a correct response ("correct rejection"/CR). D: Reaction time at test for old items (high- and low confidence hits, misses) and new items (high- and low confidence correct rejections, false alarms), by stimulus type.

Recognition Accuracy and Bias. An accurate way of measuring recognition memory and bias is to examine ROC curves, which can be constructed by plotting pairs of hit rates and false alarm rates for given levels of the participant's confidence judgment (Grider & Malmberg, 2008). However, although our experimental paradigm incorporated confidence ratings, the vast majority of the participants' responses were of high confidence (see recognition results below); in other words, several participants did not utilize the entire spectrum of confidence ratings. Therefore, to quantify recognition accuracy, we used the somewhat less accurate point measures d'=z(Hit)-z(False Alarm) as a measure of recognition sensitivity (i.e., the ease of distinguishing between an "old" and a "new" item), and c=-(z(Hit)+z(False Alarm))/2 as a measure of bias (the tendency to respond "old" vs. "new" when no information on memory strength is available; e.g. Grider & Malmberg, 2008) by collapsing over "definitely old" and "probably old" responses to obtain hit rates and false alarm rates.

Overall, sensitivity was relatively high (d'=2.27) suggesting that participants were very accurate in distinguishing between "old" and "new" items during the recognition test. The overall bias of c=-0.35 suggested that in general, participants in our experiment showed a slightly liberal bias (i.e., a relatively high probability of responding "old" when no information is available), as indexed by negative values of c.

We next performed analogous comparisons to the recall analysis: Between the three tasks, between blocks 1 and 2, 3 and 4, and 5 and 6, between sub-blocks 1 and 2, and between the three stimulus types. To do so, a slight modification in calculating of d' and c was necessary because some participants showed either a hit rate of 1 or a false alarm rate of 0 for one of the trial types, tasks or blocks. For such scenarios, neither d' nor c are

defined. Therefore, we added one false alarm and subtracted one hit for each subject before calculating d' and c. In the comparison between frequents, infrequents and pictures the transformation was more complicated because the number of picture trials was only half the number of frequents and infrequents (since pictures were only presented in the second sub-blocks). Thus, we added one false alarm and subtracted one hit for the frequents and infrequents, and added/subtracted 0.5 for pictures.

In parallel to the recall data, neither recognition sensitivity (d'; "size": M=2.12; "edible": M=2.16; "living": M=1.97) [$F(2,48)=1.65$, *ns*], nor bias (c; "size": M=-.34; "edible": M=-.26, "living": M=-.35) [$F(2,48)=1.35$, *ns*] differed between the different semantic judgment tasks. Furthermore, there were no differences between blocks 1 and 2 (d': M=2.21, c: M=-.3), blocks 3 and 4 (d': M=2.03, c: M=-.3) and blocks 5 and 6 (d': M=2.02, c: M=-.35) [d': $F(2,48)=1.97$, *ns*; c: $F(2,48)=.42$, *ns*]. Finally, there was statistically only a trend for a difference in recognition sensitivity between sub-blocks 1 (d': M=2.06; c=-.36) and sub-blocks 2 (d': M=2.17; c=-.29) [d': $t(24)=2$, $p=.06$], and the sub-blocks did not differ in bias [$t(24)=1.62$, *ns*].

However, both sensitivity [$F(2,48)=62.35$, $p<.01$] and bias [$F(2,48)=21.06$, $p<.01$] differed between the three stimulus types (figure 3B). Participants were least accurate in recognizing stimuli that were presented as frequents in the encoding phase (M=1.68), with significant differences to the infrequents (M=2.34) [$t(24)=6.92$, $p<.01$] and the pictures (M=2.95) [$t(24)=10.27$, $p<.01$]. The difference between pictures and infrequents was also significant [$t(24)=5.07$, $p<.01$].

Recognition bias (as measured by c; figure 3B) was much more conservative (i.e., participants were less likely to respond "old" without additional information) for pictures

(M=.01) than for frequents (M=-.49) [$t(24)=7.04$, $p<.01$] and infrequents (M=-.24) [$t(24)=2.61$, $p=.02$]. The difference in bias between frequents and infrequents was also significant [$t(24)=4.14$, $p<.01$].

Due to the patterns within the recall- and recognition memory analyses, we subsequently collapsed our memory measures across task, block and sub-block, but not across stimulus type.

For the purpose of an ERP subsequent memory analysis it is also important to consider the proportion of trials in each recognition category for previously studied items (high confidence hits, low confidence hits, and miss), as such proportions cannot be deduced from merely inspecting d' or c, respectively. While recognition responses to "new" events (high confidence correct rejections, low confidence correct rejections, false alarms) are not directly relevant for the purpose of an ERP subsequent memory analysis, they will be reported for the sake of completeness.

As is clearly visible in figure 3C, the proportion of previously studied stimuli that were correctly identified as "old" (i.e., the hit rate) was very high. There were no overall differences in the proportion of hits between frequents (M=.91), infrequents (M=.93) and pictures (M=.91) [$F(2,48)=.33$, *ns*]. Even when only *confident* hits were analyzed, the effect of stimulus type was non-significant [$F(2,48)=2.67$, $p=.08$], although there was a tendency for a larger proportion of high confidence hits for infrequents (M=.87) and pictures (M=.88) than for frequents (M=.82). The high hit rates obtained in our paradigm were problematic for a subsequent memory analysis because only a small number of study trials (especially for pictures and infrequents) would fulfill the "subsequently

forgotten" category (i.e., subsequent incorrect "new" responses to previously studied items; "misses").

There were larger differences between stimulus types in the proportion of previously *not* studied stimuli that were correctly identified as "new" (correct rejections; figure 3C): Both when all correct rejections [$F(2,48)=66.72$, $p<.01$], and when only confident correct rejections [$F(2,48)=80.74$, $p<.01$], were considered the differences were significant between stimulus types. The proportion of correct rejections was highest for pictures (M=.92 total, M=.81 for confident correct rejections) and lowest for frequents (M=.64 total, M=.4 only confidents), with significant differences between all pairs (all $t>4.4$). These patterns are in line with the more liberal response bias for frequents.

Overall, our data suggest that for pictures, participants were most accurate in distinguishing between previously studied and unstudied items, followed by the infrequents and, finally, the frequents. Furthermore, participants were more willing to respond "old" to frequents than to infrequents and pictures, reflected in a more liberal bias measure and a lower proportion of correct rejections.

Reaction Time during the Recognition Test. Our design yielded very high hit rates, resulting in few trials in the "forgotten" category if recognition accuracy was used to sort the study trials in an ERP subsequent memory analysis. Due to this non-optimal pattern in the recognition accuracy data, we also considered *reaction time* during the recognition test. A detailed discussion of models on reaction time distributions in recognition is beyond the scope of the present paper. However, it is important to note that most models assume that the "strength" of a memory trace is negatively correlated with reaction time during the recognition test, both for hits and correct rejections of a specific stimulus type

(e.g., Ratcliff & Murdock, 1976). Therefore, in the present paper we use recognition reaction time as an index of relative memory strength (keeping in mind that many other factors also affect recognition reaction time). An additional advantage is that reaction times provide a parametric memory measure that is suitable for a regression analysis.

Figure 3D displays the reaction times by stimulus types (frequents, infrequents and pictures) and types of recognition judgments (high- and low confidence hits, misses; high- and low confidence correct rejections and false alarms). In line with prior reports (Ratcliff & Murdock, 1976), reaction times were numerically faster for low confidence responses than for high confidence responses. Due to the high recognition accuracy, for some trial types data were not available for all participants. Therefore, we conducted the inferential statistics only on the median reaction times for high confidence hits and high confidence correct rejections. The 2 (response type: hit/correct rejection) by 3 (trial type: frequent/infrequent/picture) repeated measures ANOVA resulted in a main effect for response type [$F(1,24)=20.05$, $p<.01$], a main effect for trial type [$F(2,48)=13.18$, $p<.01$], as well as an interaction [$F(2,48)=8.72$, $p<.01$]. Thus, for all trial types, hits were faster than correct rejections (all $t>2.67$), a common finding in recognition reaction times (e.g., Ratcliff & Murdock, 1976).

Follow-up analyses conducted on confident hits and correct rejections separately revealed reaction time differences between stimulus types for correct rejections [$F(2,48)=12.65$, $p<.01$], with faster reaction times for pictures (M=922ms) than frequents (M=1166ms) [$t(24)=3.78$, $p<.01$] and infrequents (M=1081ms) [$t(24)=5.71$, $p<.01$], but no significant differences between frequents and infrequents [$t(24)=1.75$, $p=.09$]. For hits,

differences between trial types were smaller (frequents: M=806ms; pictures: M=778ms; infrequents: M=829ms), and statistically non-significant [$F(2,48)=2.84$, $p=.07$].

Overall, the faster reaction times for correct rejections, as well as the tendency for faster hits, support the idea that the memory traces were strongest for pictures, followed by infrequents, and weakest for frequents. This interpretation is in line with our recall- and recognition accuracy data and also with prior literature: The "picture superiority effect" in recall (Paivio, Rogers, & Smythe, 1968) and recognition (Shepard, 1967) is a well-studied phenomenon, and the superiority of infrequents over frequents is also in line with prior studies (e.g., Von Restorff, 1933).

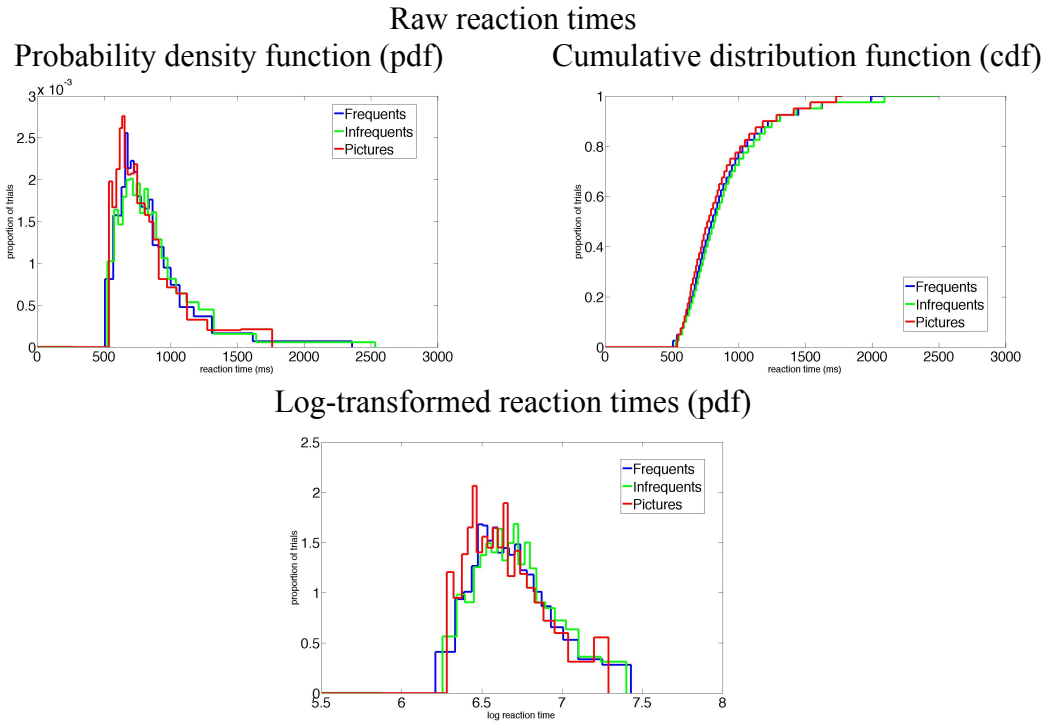


Figure 4. Distribution of reaction times for high confidence hits during the recognition test. Top panel: Vincentized probability density function (pdf) and cumulative distribution function (cdf) on the raw reaction times. Bottom panel: Log-transformed pdfs.

In a final analysis step we characterized the shape of the reaction time *distribution*, focusing on reaction times for confident hits (figure 4). Similarly to the reaction times at encoding, the distribution was positively skewed. The log-transformed distributions (after outliers deviating by more than 3 SD from the subject mean) are shown in figure 4 (bottom panel). It is important to note that although this distribution shows a much weaker positive skew than the raw reaction time distributions and that therefore the distribution more closely resembles a normal distribution, the skew is still visible.

Event-Related Potentials

In this section, we describe an analysis of the ERPs recorded in the encoding phase. In order to have enough trials in each ERP average to obtain a clean principal component analysis (PCA) solution, we collapsed ERPs across semantic judgment tasks. Thus, both the Novelty P3 and the P300 have large amplitudes and rather broad temporal distributions. Therefore, even if their latencies varied with the semantic judgment task in the same way as reaction times (the largest mean difference was 39ms between the "size" and the "edible" task), collapsing over tasks was not expected to have a major effect on the morphology of the ERPs (other than somewhat widening the average waveform) or the PCA solution. Furthermore, this effect should be the same for all conditions, so the comparisons reported below should remain unaffected.

In oddball tasks, sequential effects on ERP component amplitudes have been reported (K. C. Squires et al., 1976). To reduce the influence of such effects on our analysis as much as possible, ERPs for the frequent category included only those frequent trials that were preceded by frequents. Furthermore, to investigate a possible qualitative difference

in the ERP structure depending on whether or not pictures were present in the stimulus sequence, we initially computed ERPs separately for the first and the second sub-blocks.

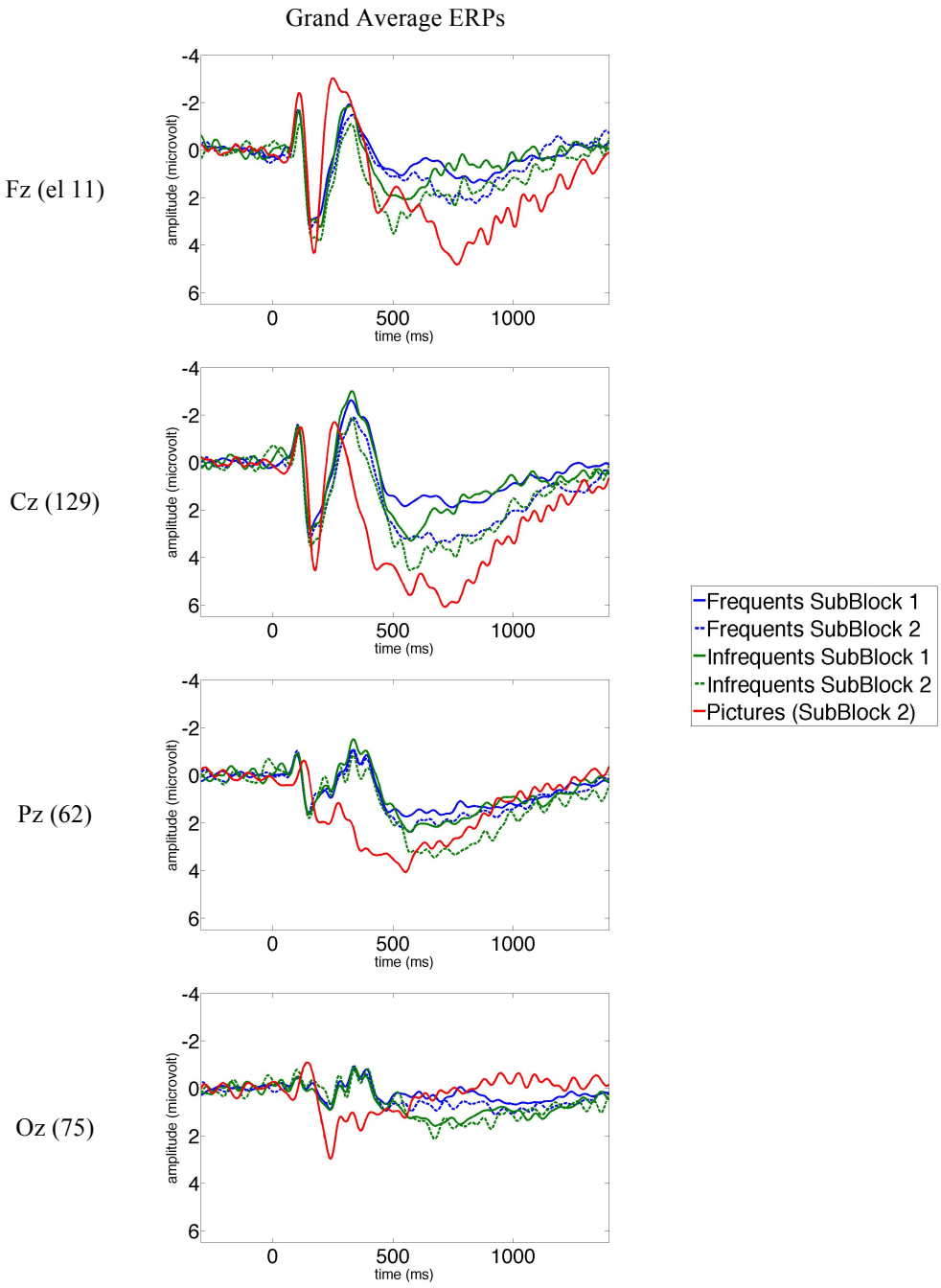


Figure 5. Grand average ERPs elicited at encoding by trial type. Displayed are frontal (Fz), central (Cz), parietal (Pz) and occipital (Oz) electrodes at midline sites.

Figure 5 shows the grand average ERPs from the frontal, central, parietal and occipital electrodes. Visual inspection suggests that the ERP patterns for frequents and infrequents were qualitatively similar for the two sub-blocks (those including pictures and those that did not). Furthermore, it is apparent that both infrequents and pictures elicited broad positivities across the scalp. A principal component analysis further characterized the componential structure of these positivities, and we specifically focused on PCA factors with spatial- and temporal distributions characteristic of the P300 and the Novelty P3.

PCA on Frequents, Infrequents and Pictures of the Two Sub-blocks. Our first PCA was conducted on the subject-averaged ERPs for five stimulus types: Frequents and infrequents in the first sub-block, frequents and infrequents of the second sub-block, and pictures (which always occurred in the second sub-block). Submitted to the PCA were recordings from all electrodes and a time window between stimulus onset (0ms) and 1400ms after the stimulus. Twenty-five spatial factors were extracted, accounting for a total of 93% of the variance, of which data from the first six factors are presented in figure 6. Based on their spatial distributions and the patterns in the virtual ERPs, the centrally distributed spatial factor 2 (variance accounted for: 26%) and the posterior factor 4 (variance accounted for: 5%) were of interest for the present study. The former may correspond to the Novelty P3 and the latter may correspond to the P300.

Importantly, infrequents and frequents in the first and second sub-blocks showed the same qualitative pattern in the ERPs: infrequents in both sub-blocks elicited both a larger Novelty P3 (spatial factor 2) and a larger P300 (spatial factor 4) than the frequents within the same sub-blocks. Furthermore, the latencies of these components did not appear to

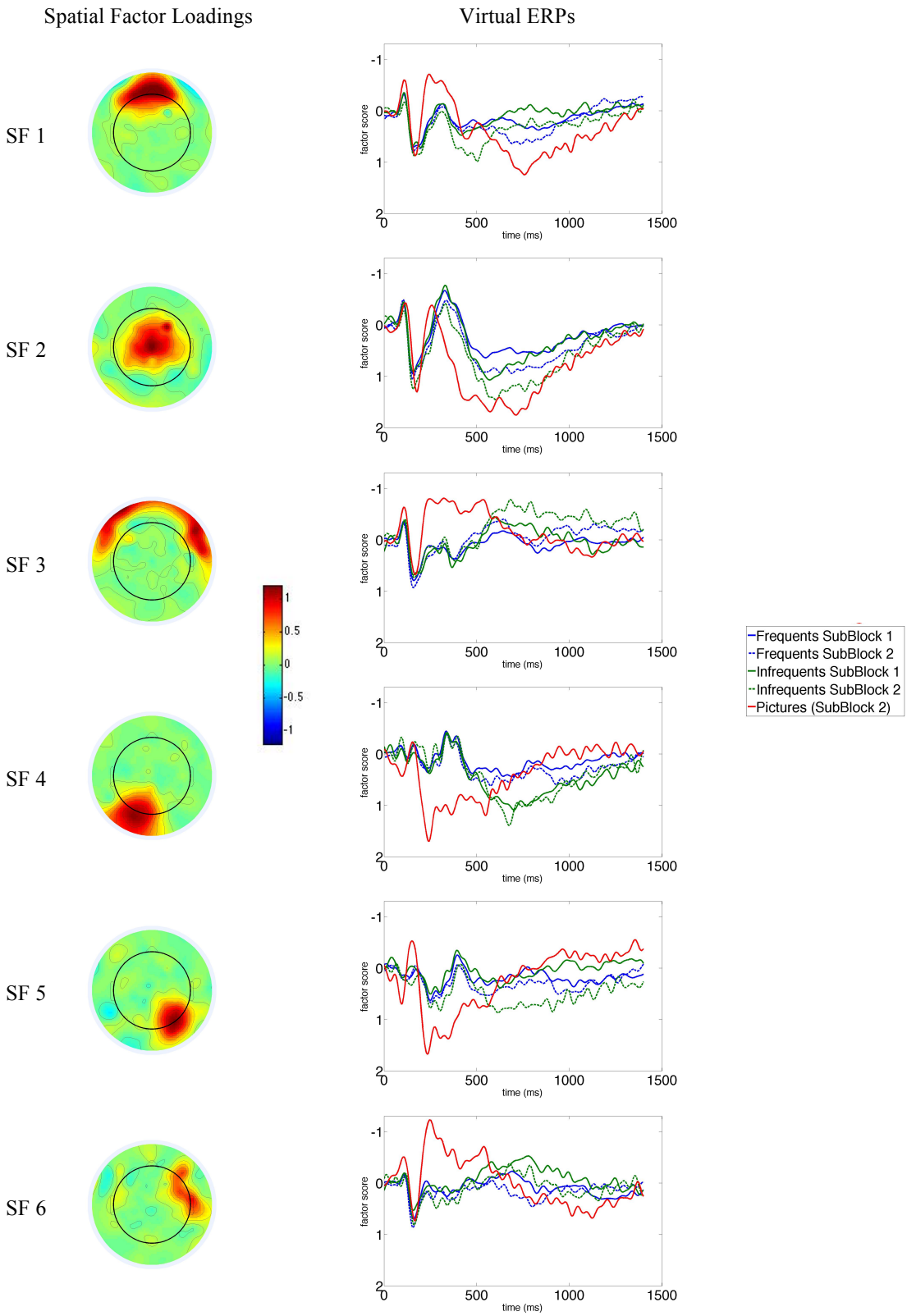


Figure 6. Results from the first PCA. Shown are spatial factor loadings (left panel), and virtual ERPs (right panel) for the first six spatial factors. *Note:* SF=spatial factor.

differ between the first and second sub-blocks. Therefore, for the purpose of simplicity we collapsed the ERP averages across sub-blocks and conducted a new PCA on this simplified dataset.

PCA on Frequents, Infrequents and Pictures Collapsed over Sub-Blocks. ERPs elicited by frequents, infrequents and pictures (collapsed across sub-blocks) were submitted to a PCA with identical parameters to the PCA described above. The solution accounted for 95% of the variance in the data. Figure 7A shows the spatial factor distributions of the first eight spatial factors. Like in the initial PCA, we obtained a central factor (spatial factor 2, accounting for 25% of the variance) and a posterior P300 factor (spatial factor 5, accounting for 4% of the variance). The distribution of their spatial factor loadings was virtually identical to the previous PCA solution. Figure 7C (left panel) shows the virtual ERPs associated with the two spatial factors of interest. Again, infrequents appeared to elicit both a larger Novelty P3 and a larger P300 than frequents. Pictures appeared to elicit the largest Novelty P3 and an earlier-peaking positivity within the P300 factor.

At this point it is worth noting that the PCA algorithm aims to extract factors that account for the highest possible percent of variance within the data. Since the pictures elicited the largest amplitudes in the relevant spatial factors, it stands to reason that the PCA solution was driven by the ERPs elicited by the pictures, therefore possibly not accurately representing the ERPs elicited by frequents and infrequents. If this is the case, then a PCA on only the ERPs elicited by verbal stimuli might yield qualitatively different factors. For this reason we finally conducted a PCA only on the ERPs elicited by the verbal stimuli (frequents and infrequents).

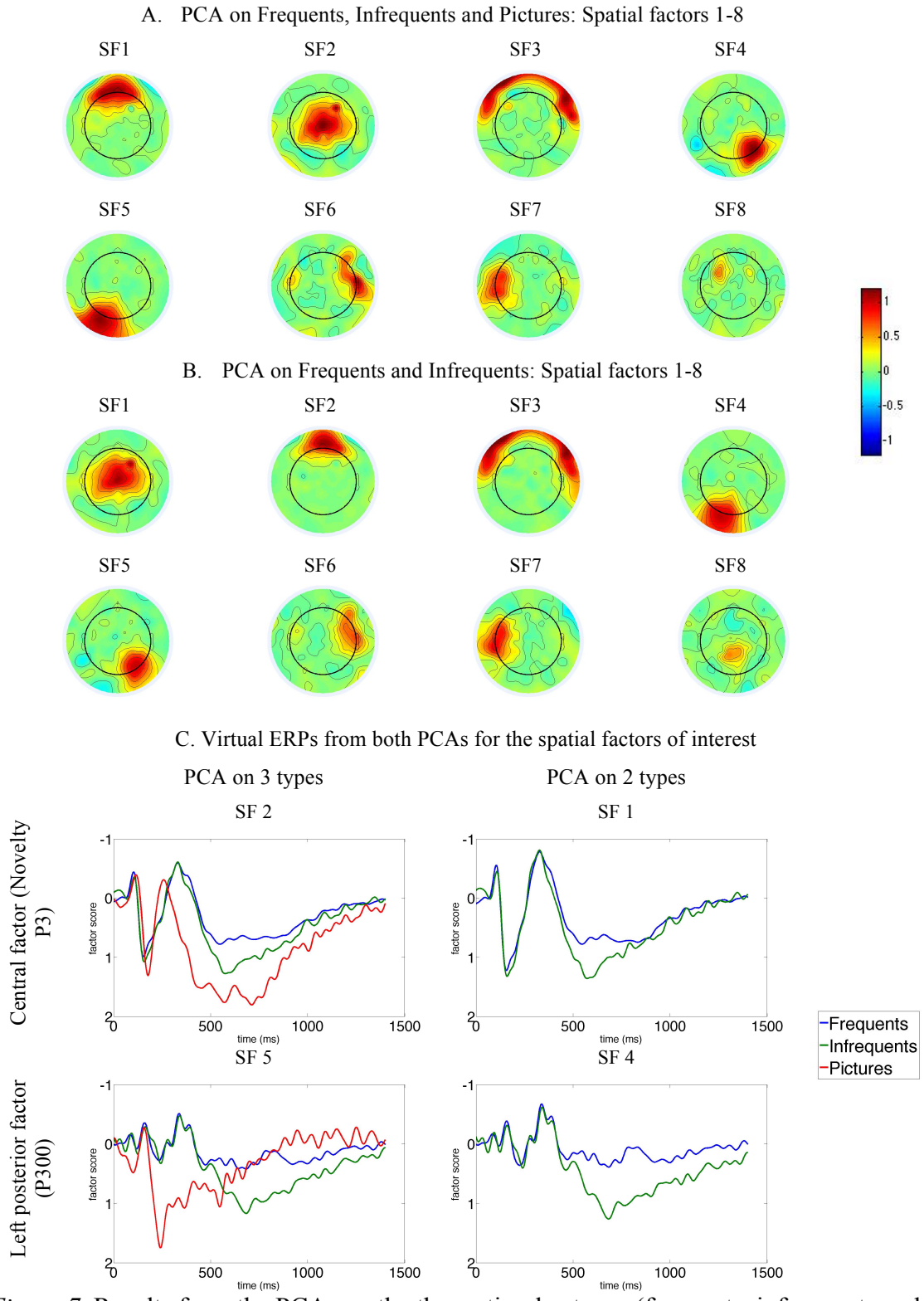


Figure 7. Results from the PCAs on the three stimulus types (frequents, infrequents and pictures), and from the PCA on two stimulus types (frequents and infrequents). A and B: spatial factor loadings of the first 8 spatial factors for the first (A) and the second (B) PCA. C: Virtual ERPs for the Novelty P3 factor and the P300 factor in each PCA.

The PCA solution accounted for 95% of the variance, and the distributions of the first eight spatial factors are presented in figure 7B. The distribution of the relevant spatial factors – SF1 as the Novelty P3 (accounting for 28% of the variance) and SF4 as the P300 (accounting for 5% of the variance), was virtually identical to the analogous factors obtained from both of the previous PCA solutions. The similarity is confirmed by an inspection the virtual ERPs, presented in figure 7C, right panel. We concluded that the spatial factors of the PCA conducted on all three trial types accurately characterize the Novelty P3 and the P300 factor for all three stimulus types.

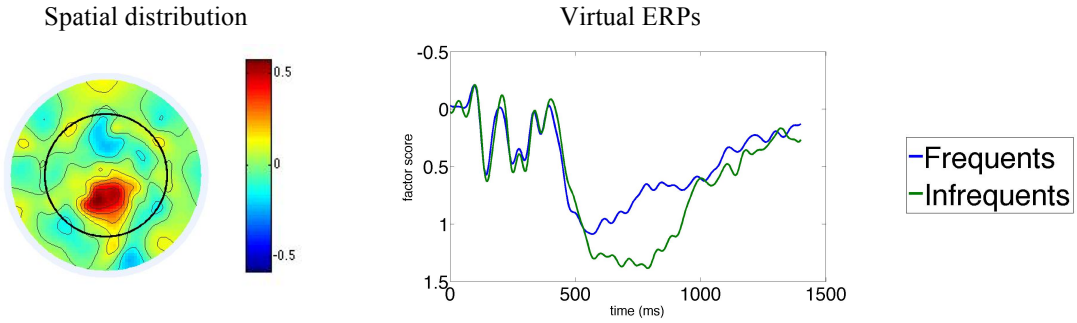


Figure 8. Additional spatial factor with the morphology of a P300. Left panel: spatial factor loadings, right panel: virtual ERPs.

It is also worth noting that the PCA on the ERPs of only the verbal stimuli revealed an additional spatial factor that exhibited the typical morphology of the P300 (figure 7B, SF8). As seen in figure 8, in the respective spatial factor's (SF 8, accounting for 3% of the variance) virtual ERPs, infrequents indeed elicited a larger positivity than frequents in the P300 time range (500-900ms after the stimulus). It is possible that this additional P300-like factor is a by-product of the PCA algorithm and that some residual variance due to the same scalp-recorded ERP component was captured in this factor as was captured by the more posterior factor obtained in both PCA solutions. However, to

account for the possibility that it represents a distinct ERP component, we included this factor in our subsequent analyses. We reasoned that if it turned out that the two spatial factors showed the same patterns in our analyses, it is likely that the two factors captured variance due to the same ERP component.

In summary, since it was clear that the PCA on the three stimulus types provided P300 and Novelty P3 factors that were representative of all stimulus types, we included the respective spatial factors in all further analyses. The second spatial factor with P300-like morphology, obtained in the second PCA, was also included as a variable.

Variance of ERPs with Trial Type. Thus far we have only visually inspected patterns in the virtual ERPs, but have not reported inferential statistics. To quantify ERP component amplitudes, the spatial PCA is typically followed by a second PCA step that identifies temporal patterns in the virtual ERPs (Spencer et al., 1999). However, temporal PCA cannot measure slight *latency differences* between conditions or individual trials and since several of our research questions addressed latency, we did not perform a temporal PCA. Instead, we used a peak picking procedure to quantify component amplitude and latency within each spatial factor of interest. The time windows used for peak picking were centered around the peak of the grand average virtual ERPs.

For the Novelty P3, we quantified component amplitude by picking the maximum spatial factor score (i.e., the maximum point in an individual's baseline-corrected virtual ERP) in a time window of 400 to 800ms after stimulus onset. Novelty P3 amplitude differed between all stimulus types [$F(2,48)=36.6$, $p<.01$], with significant differences between all pairs of stimulus types in post hoc tests [frequent vs. infrequent: $t(24)=4.07$, $p<.01$; frequent vs. picture: $t(24)=7.31$, $p<.01$; infrequent vs. picture: $t(24)=5.21$, $p<.01$].

Thus, pictures elicited the largest amplitude, followed by infrequents, and frequents elicited the smallest amplitude.

For the P300 from the first PCA, for frequents and infrequents we quantified amplitude by the maximum factor score in a time window of 500 to 900ms after stimulus onset. For the pictures, the time course of the virtual ERPs in the P300 factor was different from the other two stimulus types (figure 7C). Therefore, for the pictures we used a time window of 200 to 400 ms after the stimulus. While we still performed analogous inferential statistics, it must be kept in mind that the positivity elicited by pictures might not be the same ERP component as elicited by infrequents (see discussion). The amplitude differences between stimulus types were significant [$F(2,48)=16.21$, $p<.01$], with significant differences between all pairs of stimulus types [frequent vs. infrequent: $t(24)=3.73$, $p<.01$; frequent vs. picture: $t(24)= 5.49$, $p<.01$; infrequent vs. picture: $t(24)= 2.29$, $p=.03$]. Like for the Novelty P3, P300 amplitudes were larger for infrequents than for frequents, and were largest for pictures (again keeping in mind that the positivity elicited by the pictures might be the same as a P300).

Finally, for the additional P300-like factor obtained in the PCA on 2 stimulus types, we quantified amplitude as the maximum factor score in a time window of 500 to 900ms after stimulus onset. The visual impression of a larger P300 amplitude for infrequents than for frequents was statistically confirmed [$t(24)=3.5$, $p<.01$]. Although ERPs elicited by pictures were not included in the PCA from which this factor emerged, for the purpose of completeness we also calculated amplitude and latency measures for picture trials. Like infrequents, pictures exhibited significantly larger maximum factor scores for this

spatial factor than frequents [$t(24)=4.19$, $p<.01$]. However, there was not a significant difference between infrequents and pictures [$t(24)=1.58$, $p=.13$].

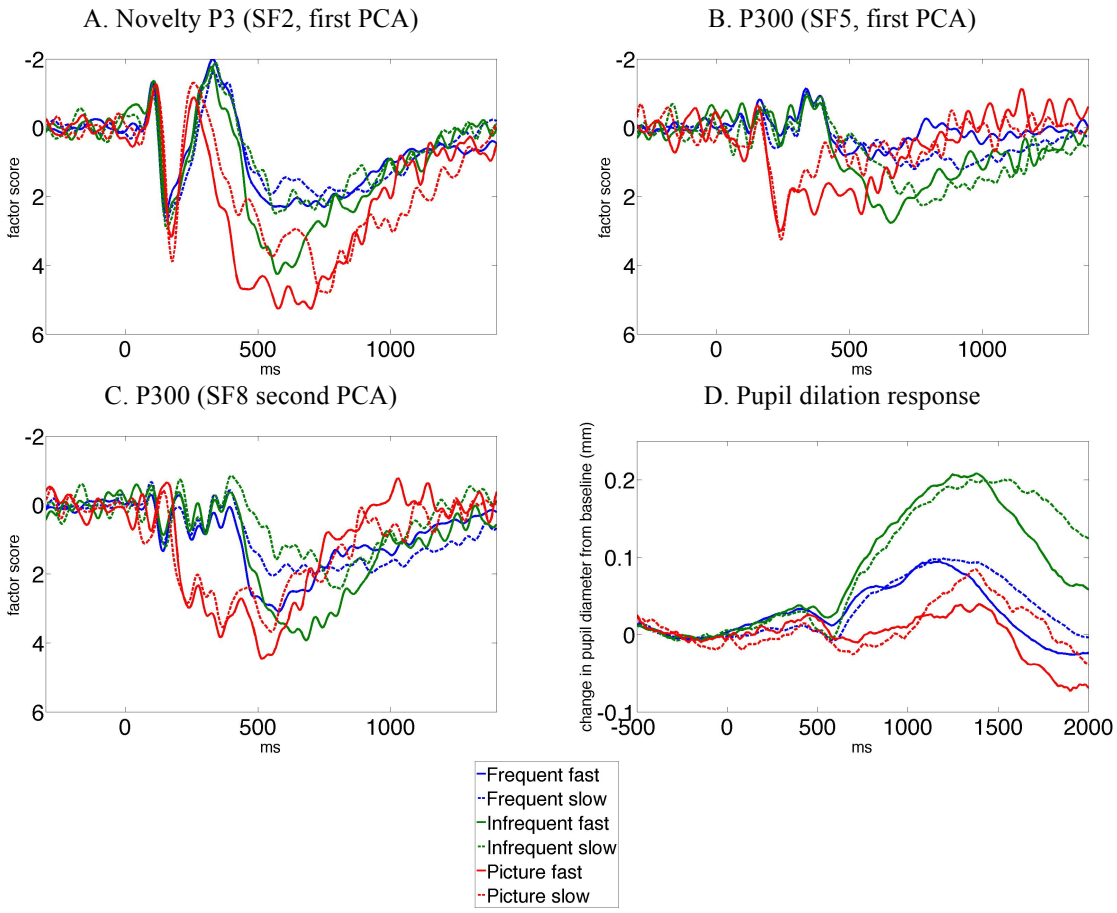


Figure 9. Relationship between ERP- and pupil measures, and reaction time at encoding. "Fast" ERPs contain trials in which reaction time was below the individual participant's median reaction time for that trial type, and "slow" responses are trials in which reaction time was above the median.

Correlation of ERP Components with Reaction Time on the Same Trial – a Median Split Analysis. Our ultimate goal was an individual trial analysis of correlations between physiological measures with behavioral measures associated with the same trial. In an initial step, we analyzed such correlations in a more traditional manner that utilized

ERP signal averaging. Thus, to analyze correlations between ERPs and reaction time at encoding, we performed a median split of reaction times for each participant and each trial type and calculated separate ERPs for "fast" and "slow" response trials. Figure 9 (A-C) shows the resulting grand average virtual ERPs for the three spatial factors of interest. We statistically analyzed the data with 3 (trial type: frequent vs. infrequent vs. picture) by 2 (fast vs. slow) repeated measures ANOVAs. Since we have already reported the variance of ERP components with trial types (frequent vs. infrequent vs. picture), we will not report main effects for trial type in this section, but will focus on main effects for speed and, wherever statistically significant or relevant for our hypotheses, interactions between speed and trial type.

As clearly visible in figure 9A, Novelty P3 amplitude was larger in trials associated with fast responses, compared to slow responses [$F(1,24)=12.08$, $p<.01$]. Response speed did not interact with trial type ($p>.46$), and the amplitude pattern was the same for all trial types (figure 9A). Novelty P3 *latency* was also correlated with reaction time on the same trial [$F(1,24)=9.05$, $p<.01$], with shorter Novelty P3 latencies for faster responses. The interaction with trial type approached significance [$F(2,48)=3.16$, $p=.05$], suggesting that the latency effect was only present for frequents [$t(24)=2.8$, $p=.01$] and pictures [$t(24)=2.47$, $p=.02$], but non-significant for infrequents [$t(24)=.13$, *ns*].

The P300 factor obtained in the first PCA did not differ in amplitude between fast and slow responses, and response speed did not interact with trial type (both $p>.6$). Instead, P300 *latency* was correlated with response speed [$F(1,24)=9.8$, $p<.01$], with shorter latencies for trials with faster responses. The interaction with trial type was also significant [$F(2,48)=3.42$, $p=.04$], suggesting that the latency effect was present only for

frequents [$t(24)=2.73$, $p=.01$] and infrequents [$t(24)=2.64$, $p=.01$], but not for pictures [$t(24)=.13$, *ns*]. This pattern is clearly visible in figure 9B.

Finally, the P300 factor obtained in the second PCA also did not exhibit significant amplitude differences between fast and slow response trials (although visual inspection of the waveforms suggest somewhat larger amplitudes for faster responses, figure 9C), and speed did not interact with trial type (both $p>.25$). However, component latency differed between fast and slow trials [$F(1,24)=5.65$, $p=.03$], with shorter latencies for trials with faster responses. The interaction was not significant [$F(2,48)=.31$, *ns*], but due to the P300 morphology differences between verbal stimuli and pictures, we still performed planned comparisons between fast and slow responses of each trial type separately. The correlation of P300 latency with reaction time was significant for infrequents [$t(24)=3.14$, $p<.01$], approached significance for frequents [$t(24)=1.73$, $p=.1$], and was non-significant for pictures [$t(24)=.65$, *ns*].

The correlations between P300 amplitude and reaction time (in both PCA factors that show characteristics of the P300) replicate a large number of prior findings (e.g., Kutas et al., 1977), in line with the idea that both reaction time and P300 latency depend on stimulus evaluation time. The correlations between Novelty P3 amplitude and latency to reaction time, in turn, are novel findings within the literature.

Correlation of ERP Components with Subsequent Recall. We next conducted a traditional subsequent memory analysis comparing ERPs elicited by subsequently recalled trials to unrecalled trials (figure 10A-C). Only 19 participants were included in this analysis because 6 participants had less than three artifact-free trials in at least one of

the ERP averages. Statistical comparisons were analogous to the median split reaction time analysis.

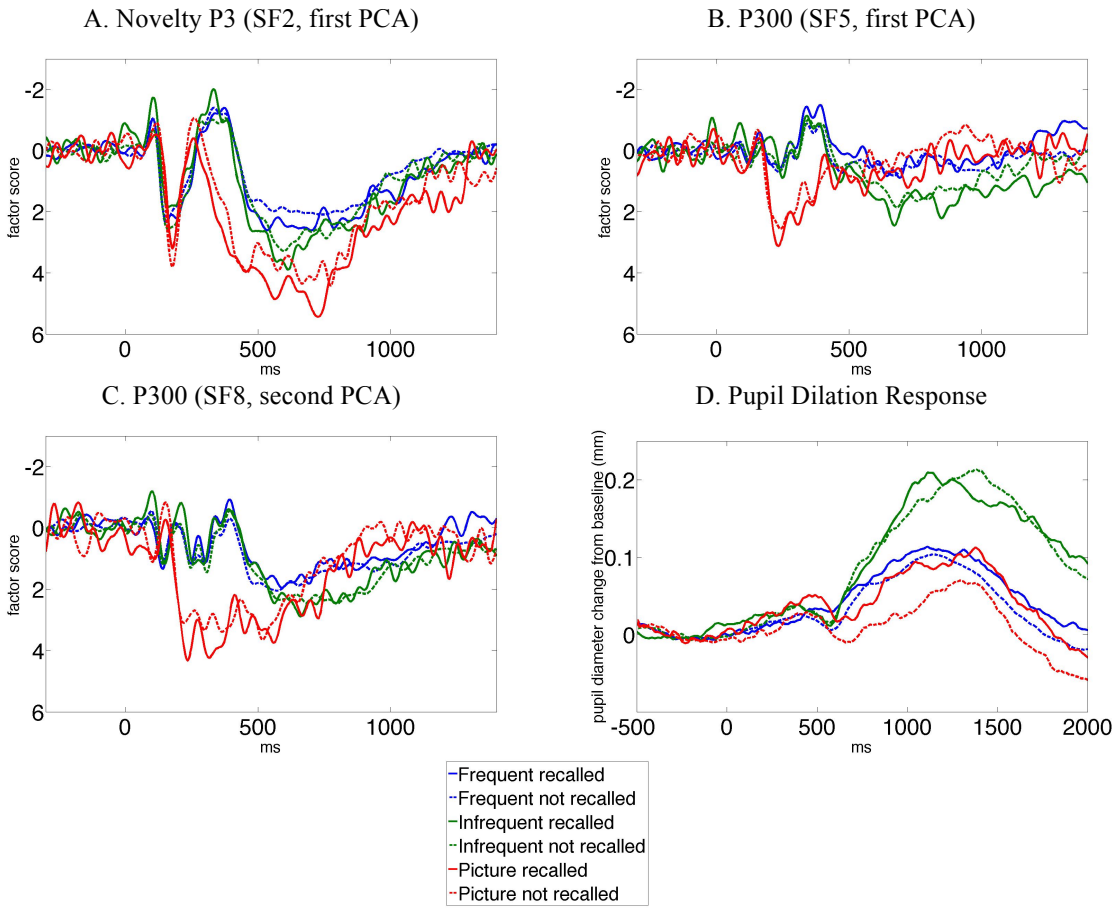


Figure 10. Grand average virtual ERPs (A-C) and pupil measures (D) elicited by subsequently recalled vs. not recalled stimuli.

Figure 10A suggests that Novelty P3 was slightly larger for later recalled, compared to forgotten, trials across stimulus types. This impression was statistically confirmed [$F(1,18)=22.86$, $p<.01$], and the difference was present for each trial type (recall by trial type interaction: $p>.98$). Novelty P3 *latency* was uncorrelated with subsequent recall and there was no interaction with trial type (both $p>.77$).

The P300 factor obtained in the first PCA also exhibited larger amplitudes for subsequently recalled stimuli [$F(1,18)=11.77$, $p<.01$], although inspection of figure 10B suggests that the differences were small. In planned comparisons, this subsequent memory effect was significant for infrequents [$t(18)=2.32$, $p=.03$] and pictures [$t(18)=2.12$, $p<.05$], but not for frequents [$t(18)=1.67$, *ns*]. P300 *latency* in this spatial factor was not correlated to subsequent recall [$F(1,18)=.57$, *ns*].

Finally, the P300 factor obtained in the second PCA also showed larger amplitudes for subsequently recalled than not recalled trials [$F(1,18)=11.28$, $p<.01$]. The interaction was not significant ($p>.88$). We again performed planned comparisons on each trial type, but the subsequent memory effect was not robust and thus statistically non-significant for each trial type alone ($p>.09$). P300 latency in this spatial factor was uncorrelated with recall [$F(1,18)<.01$, *ns*].

The P300 subsequent memory effect we report in this section is in line with many previous studies (e.g., Karis et al., 1984), but the effect actually appeared to be smaller in magnitude than the Novelty P3 effect in the present data set. The Novelty P3 subsequent memory effect is a relatively novel finding, but in line with this pattern, a study from our group recently reported a correlation between the Novelty P3 (where the spatial distribution of the Novelty P3 was similar to the present data set) and subsequent recall (Kamp et al., in press).

Correlation of ERP Components with Reaction Time at Test. Since the proportion of high confidence hits was too high, not enough trials would have been in the "forgotten" category of a traditional subsequent memory analysis using recognition accuracy as the means of sorting ERPs. Therefore, we focused on only subsequent

confident hits and sorted encoding trials based on subsequent reaction time during the recognition test, as an index of memory strength. Like for the reaction time at encoding, we performed a median split and compared "subsequent fast confident hits" to "subsequent slow confident hits" (figure 11). One participant was excluded due to insufficient trial numbers, resulting in 24 participants for this comparison.

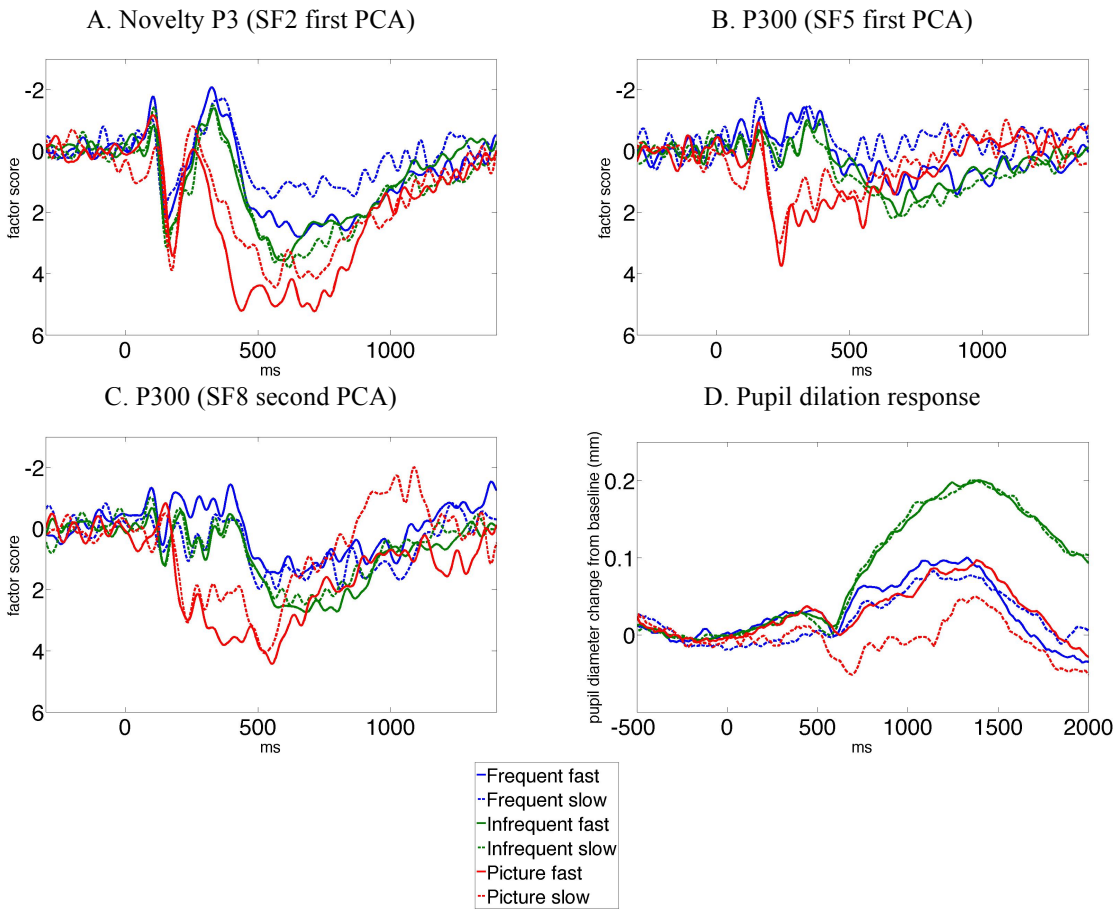


Figure 11. Relationship between ERP- and pupil measures, and reaction time at test. "Fast" ERPs contain trials in which reaction time was below the individual participant's median reaction time at test for that trial type, and "slow" responses are trials in which reaction time was above the median.

Visual inspection of figure 11A suggests that larger Novelty P3 amplitudes were associated with faster subsequent recognition judgments (at least for frequents and pictures). However, the main effect for subsequent response speed was neither statistically significant for Novelty P3 amplitude [$F(1,23)=.38$, *ns*], nor for Novelty P3 latency [$F(1,23)=.79$, *ns*]. Similarly, the P300 from the first PCA (figure 11B) did not show significant main effects of subsequent recognition speed in its amplitude [$F(1,23)=.05$, *ns*] or latency [$F(1,18)=.23$, *ns*]. We also performed planned comparisons between fast and slow subsequent recognition judgments for each trial type, but none revealed any statistically significant, systematic variance with subsequent recognition reaction time.

The amplitude of the P300 obtained in the second PCA exhibited a main effect for subsequent recognition reaction time [$F(1,23)=5.72$, $p=.03$] – this effect, however, was in the opposite direction as predicted: larger amplitudes were associated with slower responses at test. The interaction with stimulus type also approached significance [$F(2,46)=2.55$, $p=.09$], indicating that larger amplitudes for trials that were associated with slower subsequent reaction times during the recognition test - was only significant for frequents [$t(18)=2.57$, $p=.02$] ($p>.39$ for the other trial types).

Overall, besides some interesting trends, the median split ERP analysis using the reaction time during the recognition test revealed weak or no statistically robust differences in any of our measures between fast and slow recognition judgments. However, reaction time during recognition provides a continuous measure that can index subsequent memory strength and is therefore well suited as a dependent variable in a

regression analysis. A regression analysis on single trials may also have a larger power to detect a correlation between physiological measures and reaction time at test.

Correlation of ERP Components with Reaction Time on the Next Trial. Previous studies (Notebaert et al., 2009) have reported that after the presentation of infrequent stimuli, responses to the next stimulus are slowed down. Furthermore, the amplitude of the P300 has been shown to correlate with the extent to which reaction times are slowed down to the next trial after an error (e.g., Hajcak et al., 2003). These prior findings motivated us to study the extent to which amplitudes and latencies of the physiological measures elicited in our Novelty P3 oddball paradigm were correlated with reaction time to the next trial. It is worth noting, however, that our behavioral data exhibited only a statistically non-significant trend for slowed responses after infrequent stimuli, suggesting that in our dataset the power of detecting associations between physiological measures and reaction time to the next trial might also be reduced.

The analysis compared trials followed by a fast response to trials followed by a slow response, using a median split, given that the following trial was of the frequent category (figure 12A-C). Two participants were excluded from this analysis due to insufficient trial numbers.

Novelty P3 amplitude [$F(1,22)<.01$, *ns*] and latency [$F(1,22)=.17$, *ns*] were unrelated to reaction time on the next trial (figure 12A).

The P300 from the first PCA (figure 12B) showed no amplitude differences based on subsequent reaction time [$F(1,22)=1.13$, *ns*], but its latency varied with reaction time to the next trial [$F(1,22)=10.99$, $p<.01$], with the interaction approaching significance [$F(2,44)=2.89$, $p=.07$]. Thus, shorter P300 latencies were associated with subsequent

faster responses only for frequent [$t(22)=2.61$, $p=.02$] and infrequent [$t(22)=2.57$, $p<.01$] trials, but not for pictures [$t(18)=.15$, *ns*].

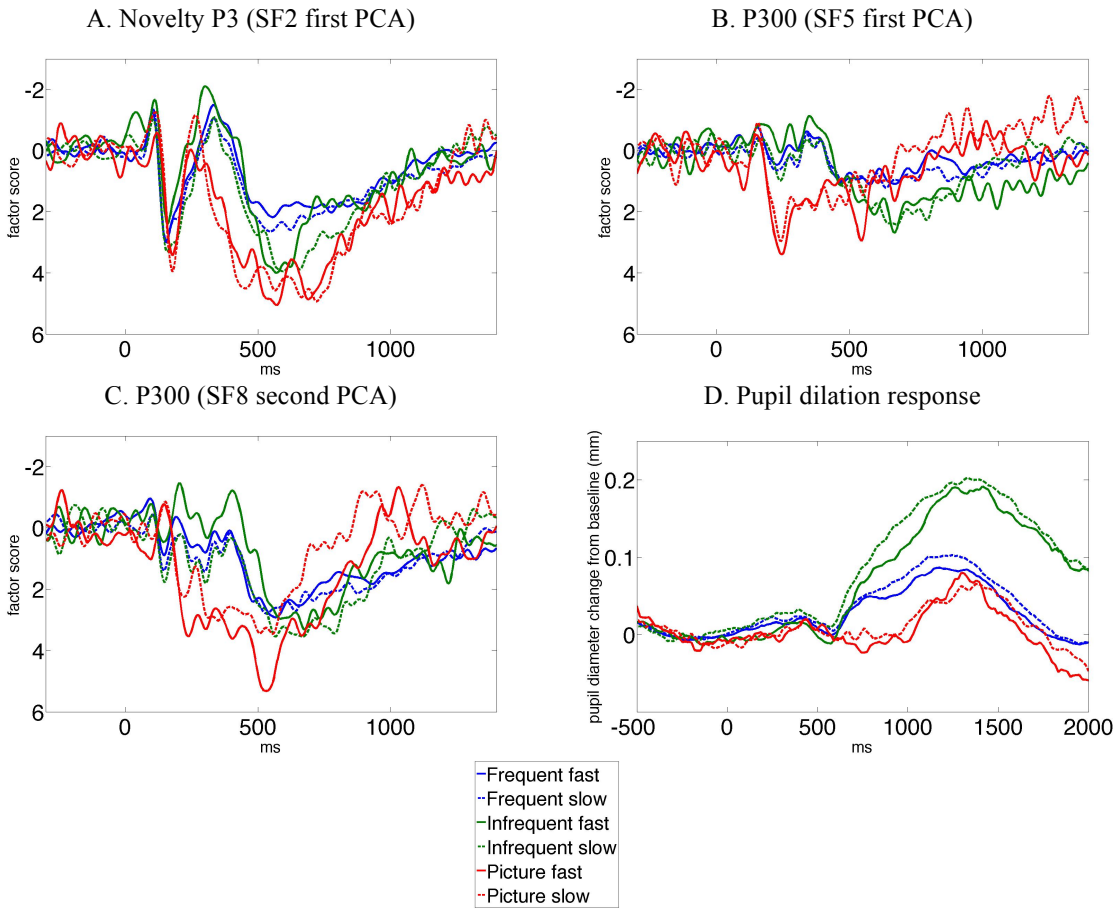


Figure 12. Relationship between ERP- and pupil measures, and reaction time to the immediately following trial. Included are only trials followed by frequents. "Fast" ERPs contain trials in which reaction time to the next trial was below the individual participant's median reaction time for frequents, and "slow" responses are trials in which reaction time was above the median.

For the amplitude of the P300 from the second PCA (figure 12C) the effect of subsequent response speed was non-significant [$F(1,22)=2.78$, $p=.11$], and for its latency the main effect only approached significance [$F(1,22)=3.59$, $p=.07$]. In subsequent

planned comparisons, the association between shorter P300 latencies and faster subsequent responses was significant only for infrequent trials [$t(22)=2.61$, $p=.02$], but not for frequents [$t(22)=1.68$, *ns*] or pictures [$t(22)=.22$, *ns*].

Similarly to the recognition reaction time analysis, the effects in the median split analysis using reaction time to the next trial were weak and not generally statistically robust. However, reaction time to the next trial was also analyzed in a regression model, which might provide greater power to detect differences.

Pupil Data

In this section we report the analysis of our pupil size measures recorded during the encoding phase. Seven of the twenty-five participants were excluded from the pupillometric analysis due to equipment failure of the eye tracker or human error collecting the eye tracking data (n=3) or due to excessive noise (such as due to eye blinks) in the pupil size recording (n=4).

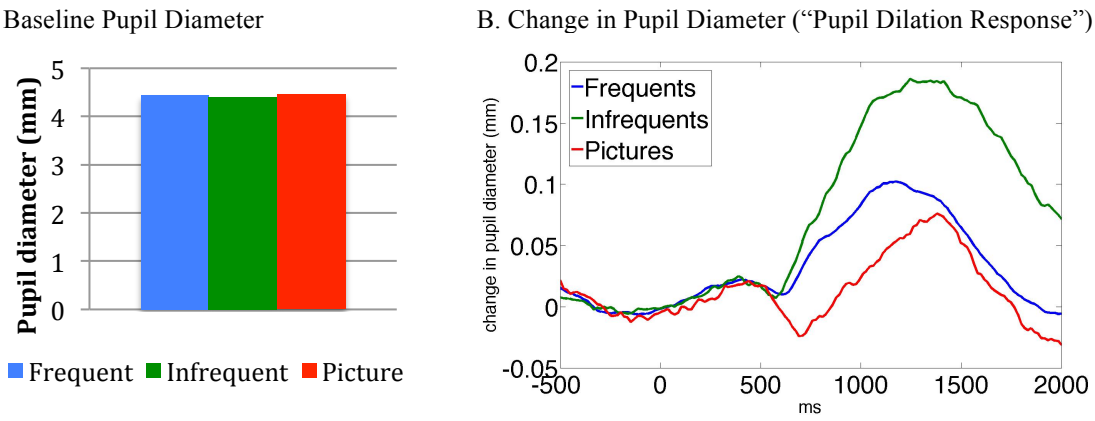


Figure 13. Differences in pupil measures between stimulus types. A: Baseline pupil diameter for frequents, infrequents and pictures. B: Temporal progression of changes in pupil diameter from baseline, time-locked to stimulus onset.

In line with the ERP analysis we collapsed trials across the first and second sub-blocks. Figure 13A shows the average baseline pupil diameter recorded over the 500ms preceding each stimulus. Since during the baseline period it was unknown which stimulus type would next be presented, it is not surprising that the baseline pupil diameter was comparable between trials with the three different stimulus types (about 4.5mm). Figure 13B shows the change in pupil diameter from baseline over time, time-locked to stimulus onset. Thus, infrequents appeared to elicit a larger increase in pupil size than frequents – a change of about 0.2mm, compared to 0.1mm. Pictures initially elicited a small "dip" in the pupil diameter recording, possibly a light reflex to the offset of picture presentation (recall that stimuli stayed on the screen for 300ms), followed by a dilation. However, the magnitude of this dilation was overall numerically *smaller* than for frequents.

We first analyzed the maximum amplitude of the pupil dilation response (PDR) for each stimulus type, using a time window of 1000ms to 1500ms after stimulus onset for peak-picking (figure 13B). In line with the visual impression, there were differences between stimulus types in PDR amplitude [$F(2,34)=27.95$, $p<.01$]: Greater dilations were elicited by infrequent stimuli than by both frequents [$t(17)=6.06$, $p<.01$] and pictures [$t(17)=6.53$, $p<.01$], with no significant difference between frequents and pictures [$t(17)=1.06$, *ns*]. While the latency of the peak dilation appeared to be earlier for frequents than for infrequents and pictures, this difference was not significant [$F(2,34)=2.48$, $p=.1$].

In order to analyze whether pupil size returned to baseline earlier for frequents than for infrequents (as suggested by an inspection of figure 13B), we next analyzed the mean dilation (compared to the baseline) in a second time window, 1500ms to 2000ms after

stimulus onset. Indeed, the difference between stimulus types was significant [$F(2,34)=19.92$, $p<.01$], with larger mean amplitudes for infrequents than for frequents and [$t(17)=6.64$, $p<.01$] and pictures [$t(17)=4.59$, $p<.01$], but no significant difference between frequents and pictures [$t(17)=.77$, *ns*]. This suggests that the larger amplitude of the pupil dilation response for infrequents was followed by a slower return to the baseline diameter.

Since only infrequent stimuli required an infrequent response (i.e., a response switch), one possible interpretation of the fact that only infrequents, but not pictures, elicited larger dilations that frequents, is that the pupil dilation response indexes processes related to behavioral responding. Our data are also in line with Kahneman's (1973) suggestion that pupil size indexes "effort" exerted by the participant, as a switch in response can reasonably be assumed to require more "effort" than the execution of the frequent response. Note, however, that Kahneman recorded pupil size over a longer period of time than the time window in which the phasic PDR is observed. The analysis reported next addresses in more detail the relationship between the PDR and behavioral responding.

Correlation of Pupil Measures with Reaction Time at Encoding – A Median Split Analysis. Like for the ERPs, we performed a median split on reaction time at encoding and thus calculated separate pupil averages for "fast" and "slow" response trials (figure 9D). Two participants were excluded from this analysis due to an insufficient number of trials in one of the categories. The statistical analysis was analogous to the ERP analysis.

As is visible in figure 9D, the maximum dilation did not differ between trials with fast and slow responses [$F(1,15)=.57$, *ns*]. Rather, the main difference due to response speed appeared to be within the *latency* to the peak dilation and the amplitude in the second

time window. This was corroborated by a main effect of response speed on peak latency [$F(1,15)=10.78$, $p<.01$]. This pattern was visible for all stimulus types (figure 9D), and the interaction between stimulus type and response speed was non-significant ($p>.64$). Similarly, the mean amplitude in the second time window was larger for slow- than for fast responses [$F(1,15)=14.94$, $p<.01$], a pattern that was again seen for all stimulus types (figure 9D; interaction with stimulus type $p>.74$). In other words, slower responses were associated with a later peak along with a slower return to the baseline diameter.

Baseline pupil diameter did not differ between fast and slow response trials [$F(1,15)=.87$, *ns*], although in planned comparisons there was some evidence for smaller baseline diameters in trials with faster reaction times for frequents [$t(15)=2.23$, $p=.04$]. However, comparisons of baseline diameter in fast and slow trials for the other trial types were non-significant ($p>.75$), and for infrequents the difference was even in the opposite direction (i.e., baseline diameters tended to be *larger* for trials with shorter reaction times).

Subsequent Memory Analysis: Recalled vs. Not Recalled Trials. Analogously to the ERP analysis, we also compared pupil measures between subsequently recalled- and non-recalled trials (figure 10D). Two participants were excluded from this analysis due to an insufficient number of trials.

The maximum change from baseline in the time window 1000-1500ms indeed distinguished between subsequently recalled- and not recalled items, as can be seen especially for frequents and pictures in figure 10D. This subsequent memory effect was significant overall [$F(1,15)=11.77$, $p<.01$]. However, subsequent planned comparisons between recalled and not recalled trials of each stimulus type failed to reach significance

in all cases (p>.07). The latency to the peak dilation did not show a subsequent memory effect [F(1,15)=1.16, *ns*], and the difference in dilation in the second time window only approached significance [F(1,15)=3.26, p=.09]. In subsequent planned comparisons the difference in the second time window was not robust for any trial type (p>.1).

Finally, baseline pupil diameter did not vary with subsequent recall [F(1,15)=.79, *ns*].

In summary, there was some evidence for an association between larger peak pupil dilations and subsequent recall success, but this effect did not survive planned comparisons for each stimulus type. Other than that, none of the pupil measures showed subsequent memory effects.

Median Split Analysis of Reaction Time at Recognition. We also split the encoding trials into "fast" and "slow" subsequent recognition judgments, using a median split on the associated reaction time at test (only for trials that had been subsequently given a confident "old" response; figure 11D). Data from two participants were not included due to insufficient trial numbers.

There were no significant main effects for reaction time at test, neither for the maximum dilation [F(1,15)=1.78, *ns*], nor for peak dilation latency [F(1,15)=.4, *ns*], nor for the mean amplitude in the "return to baseline" time window [F(1,15)=.14, *ns*]. In reflection of the apparent amplitude difference for pictures that received fast, compared to slow, recognition judgments, the interaction between trial type and recognition speed approached significance for the mean amplitude in the second time window [F(2,30)=3.22, p=.05]. However, none of the subsequent planned comparisons were significant (p>.1). Overall, as for the recall data, associations between pupil dilation measures and subsequent recognition were absent or weak.

Interestingly, the analysis revealed a correlation between the baseline diameter and subsequent recognition response speed. That is, generally, trials with smaller baseline diameters at encoding were associated with quicker "old" responses during recognition [$F(1,15)=10.7$, $p<.01$].

Pupil Measures and Reaction Time to the Next Trial. Figure 12D shows the pupil measures for trials that were followed by "fast", and "slow" responses, respectively. Note that three participants did not have enough trials for the analysis and were thus excluded. As can be seen in the figure, the pupil dilation response to trials followed by fast and slow subsequent responses were comparable. In corroboration of this impression, neither peak amplitude [$F(1,14)=.2$, *ns*], nor peak latency [$F(1,14)=.44$, *ns*], nor the mean amplitude in the second time window [$F(1,14)=1.62$, *ns*] differed between trials with fast- and slow subsequent reaction times. Likewise, baseline pupil diameter was uncorrelated with reaction time to the next trial [$F(1,14)=1.99$, *ns*].

Regression Models

The analyses reported so far have provided some insight into the response of each physiological measure to experimental manipulations in a Novelty Oddball task. The patterns that we discovered suggest that the sensitivity to task parameters of the pupil dilation response, the Novelty P3 and the P300 do not overlap perfectly and hence most likely, the three physiological responses reflect different cognitive processes invoked upon the encounter of novelty.

The interpretation of the patterns in the P300 was complicated by the fact that two PCA factors with the morphology of the P300 were obtained- one posterior factor obtained in the first PCA, and a parietal factor in the second PCA on only verbal stimuli.

The virtual ERPs of both factors clearly distinguished between frequents and infrequents, and pictures elicited an earlier positivity. Due to the early peak it is unlikely that this ERP component is a P300; instead, the occipital distribution may suggest that it reflects early visual processes invoked by perceptual deviance (for further discussion of this issue, see discussion section). Overall, P300 latency (in either P300 factor) was correlated with reaction time on the same trial, particularly for frequents and infrequents. Furthermore, P300 amplitude predicted subsequent recall. These patterns alone do not strongly imply a role of the P300 for either the immediate behavioral reaction or memory encoding.

The Novelty P3 was largest for perceptually salient stimuli (i.e., pictures of the frequent category inserted into the oddball stream), which required the frequent response and were also associated with the strongest memory traces, as indexed by our behavioral analysis. Combined with the fact that response time to pictures was indistinguishable from frequents, this suggests that the Novelty P3 is not directly related to response preparation or execution. Rather, it might index the processing of deviance, with a specific sensitivity to *perceptual* deviance. However, somewhat inconsistent with this idea, Novelty P3 amplitude was correlated with both reaction time on the same trial, as well as the probability of subsequent recall. One possibility is that this ERP component is sensitive more generally to resource allocation to novel stimuli.

Finally, the pupil dilation response was largest for infrequent stimuli and tended to be *smaller* in amplitude for pictures than frequents. This indicates that the PDR does not index the processing of perceptual deviance or stimulus probability *per se* (infrequents and pictures were equally probable in sub-blocks 2), but it might be sensitive to response preparation or execution demands that are heightened when an infrequent response is

required. In line with this idea, the latency of the peak dilation and its return to baseline were correlated with reaction time. However, the maximum amplitude was also correlated with subsequent recall, so thus far, as for the other measures, we are thus far unable to link the pupil response uniquely to either behavioral responding or memory.

The expectation that all measures will be correlated to some extent with both measures of immediate responding and subsequent memory was therefore confirmed by our data. Next, a sequence of regression analyses investigated the correlation between each physiological measure to behavior on an individual trial basis. This analysis allowed us to (1) test whether the relationships observed in the median-split analysis held up in an analysis focused on individual trials and (2) whether the each physiological measure remained a significant predictor of the respective behavioral measure *when the other physiological responses had been accounted for*. Note that only the 18 participants for whom pupil data were available were included in the regression analysis.

The dependent variables were reaction time at encoding, reaction time at test, and reaction time to the next trial. We did not perform an analogous analysis for the recall data, because recall is a binary variable and therefore not ideally suited for parametric regression.

Correlations between Physiological Measures and Reaction Time at Encoding. The first question was which physiological measures were correlated with reaction time on an individual trial basis. It is worth re-iterating that the physiological measures of interest were the *amplitude* and *latency* of the Novelty P3, the P300 from the first PCA on the 3 stimulus types, the P300 from the second PCA on 2 stimulus types, and the pupil dilation response, as well as the baseline pupil diameter and the mean change in pupil

diameter from baseline in the second time window (the "return to baseline" time window).

Table 1 shows partial correlation coefficients between each physiological variable and reaction time on the same trial during encoding. The first value in each cell represents the overall correlation when variance due to the variables *participant, encoding task, sub-block,* and *stimulus type* are partialed out. The next three entries represent the partial correlations computed separately for frequents, infrequents, and pictures, respectively. Note that the separate correlations for each trial type are listed in order to be able to accurately interpret the patterns and their variance with stimulus type. However, for the purpose of simplicity, the regression analyses always included data from all three stimulus type, with "stimulus type" included as a covariate.

The first column in table 1 reflects which physiological variables are individually correlated with reaction time. Overall, these correlations show striking parallels to the median split analysis of reaction times. First, Novelty P3 amplitude was inversely correlated with reaction time on the same trial: A larger Novelty P3 was associated with a faster response on the same trial. The correlation between Novelty P3 *latency* and reaction time was only significant when infrequent trials were analyzed separately.

Also in line with the median split analysis, for both the P300 (measured in either spatial factor) and for the pupil dilation response, it was not peak *amplitude*, but peak *latency* that was correlated with reaction time. That is, longer P300 latencies and longer latencies of the PDR were associated with longer reaction times. In addition, the maximum amplitude of the P300 from the second PCA also showed a negative correlation with reaction time for pictures and infrequents, but not across stimulus types.

Table 1: Correlations among log reaction time at encoding and physiological variables, when variance due to *participant, task, stimulus type* and *sub-block* has been partialed out.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Log RT (Encoding)	-										
2. Novelty P3 Max	**-.11**** ***-.1***** ***-.18***** ***-.12*****	-									
3. Novelty P3 Latency	.03 *-.00* ***.15***** *.1*	**.1**** ***.11***** *.04* *.09*	-								
4. P300 Max (1st PCA)	-.02 *-.03* *-.08* *.01*	**.39**** ***.41***** ***.38***** ***.25*****	**.06**** ***.06***** *.09* *.09*	-							
5. P300 Latency (1st PCA)	**.1**** ***.09***** ***.16***** *-.06*	-.00 *.01* *-.07* *.07*	**.11**** ***.11***** ***.18***** *.03*	**.05**** ***.05**** *.02* *.12*	-						
6. P300 Max (2nd PCA)	-.04 *-.02* ***-.11**** ***-.15****	**.34**** ***.33***** ***.36***** ***.39*****	**.06**** ***.05**** *.08* *.13*	**.46**** ***.49***** *.44*** ***.26*****	.04 *.03* *.02* *.12*	-					
7. P300 Latency (2nd PCA)	**.11**** ***.12***** ***.11**** *.03*	-.02 *-.02* *-.06* *.04*	**.13**** ***.13***** ***.15***** *.11*	.01 *.04* *-.03* *-.1*	**.34**** ***.35***** ***.37***** *.01*	.03 *.04* *-.02* *.02*	-				
8. Pupil Dilation Max	.01 *.02* *-.07* *.09*	-.03 *-.04* *.07* *-.01*	-.00 *.00* *.01* *-.06*	-.01 *.00* *-.00* *-.04*	**.04*** ***.05**** *.01* ***-.15****	-.01 *-.01* *.01* *-.00*	**.06**** ***.07***** *.01* *-.07*	-			
9. Pupil Dilation Latency	**.18**** ***.18***** ***.15***** ***.23*****	**-.04*** *-.04* *-.08* *.08*	.01 *-.01* *.06* *-.07*	**-.04*** *-.03* *-.08* *-.08*	**.05**** ***.05**** *.06* *-.13*	**-.05**** ***-.06**** *-.05* *-.05*	.02 *.03* *-.01* *-.06*	**.11**** ***.11***** ***.11**** ***.14****	-		
10. Pupil Mean 2nd TW	**.1**** ***.1***** ***.1**** ***.15****	**-.04*** *-.05* *.00* *-.03*	-.01 *-.01* *.01* *-.06*	-.04* *-.03* *-.06* *-.04*	.03 *.03* *.02* *-.09*	**-.06**** ***-.06***** *-.09* *-.01*	**.05*** ***.05**** *-.01* *.07*	**.79**** ***.78***** ***.81***** ***.81*****	**.3**** ***.3***** ***.27***** ***.31*****	-	
11. Pupil Baseline	**.1**** ***.11***** *.05* ***.18*****	-.01 *-.00* ***-.1**** *.05*	**-.04*** ***-.07***** *.03* *-.07*	-.03 *-.01* ***-.11**** *.03*	.02 *-.00* ***.1**** *.04*	-.00 *.00* *.01* *-.07*	-.02 *-.02* *.06* *-.11*	**-.39**** ***-.37***** ***-.46***** ***-.45*****	-.00 *.01* *-.06* *-.01*	**-.42**** ***-.4***** ***-.45***** ***-.45*****	-

The first number in each cell is the partial correlation coefficient between the two measures with *stimulus type* partialed out. The remaining three numbers represent the correlation coefficient for (1) frequents, (2) infrequents or (3) pictures only. Shaded cells in the first column index physiological measures correlated (across stimulus types) with log reaction time. Shaded areas in the rest of the table indicate correlation coefficients between two variables that are both individually correlated with reaction time. *Note:* ** indexes $p<.01$, * indexes $p<.05$

A final pattern that is in line with the median split analysis is the correlation between reaction time and the mean change from baseline in the second time window of the PDR (the "return to baseline" time window). Slow responses were associated with larger dilations in the second time window, suggesting a slower return of the pupil size to baseline.

An additional finding not obtained in the median split analysis was the correlation between baseline pupil diameter and reaction time: Smaller baseline amplitudes were associated with shorter reaction times.

In summary, many physiological variables were correlated with reaction time on the same trial. It is worth investigating whether the physiological responses are also correlated with *each other,* possibly due to third variables such as "attentional resources" allocated in each trial. Thus, we examined the correlations among the variables, with a particular focus on those physiological variables that were correlated with reaction time (shaded cells in table 1). Several relatively highly correlated pairs of physiological measures are worth noting. P300 latency measured by the first PCA on three stimulus types, and P300 latency measured by the second PCA showed a relatively high overall partial correlation of r=.34. Furthermore, the mean change in pupil diameter from baseline in the second time window (the "return to baseline" time window) was correlated with peak dilation latency (r=.3). This correlation is intuitive because a later peak in the dilation of the pupil should also temporally delay the return of pupil size to baseline. Finally, baseline diameter was negatively correlated with mean diameter change in the second time window. In other words, larger baseline pupil sizes were associated with a faster return to baseline. These correlations among physiological

measures must be kept in mind when interpreting the outcome of our multiple regression model because collinearity of predictor variables can under some circumstances be problematic in regression models.

Thus, we next developed a multiple regression model to predict reaction time at encoding on an individual trial basis, using those physiological measures as predictors that were individually correlated with reaction time. The goal was to investigate which of the measures will statistically predict reaction time *when the other variables have been accounted for*. The regression model included the random covariate *participant* as well as the fixed covariates (known, based on our behavioral analysis, to affect reaction time) *task type*, *sub-block*, and *trial type*.

Table 2. Multiple regression model on the log reaction times at encoding, which included *participant* as a random covariate, as well as *task*, *stimulus type* and *sub-block* as fixed covariates.

Variable	Unstandardized Coefficient (B)	Standard Error	Standardized Coefficient (β)
Novelty P3			
Amplitude	-.00168	.000306	**-.104****
P300 (PCA 1)			
Latency	.000059	.000019	**.07679****
P300 (PCA 2)			
Latency	.000069	.000018	**.06659****
Pupil			
Latency	.000105	.000014	**.1325****
Mean dilation 1.5-2s	.05342	.01018	**.1004****
Baseline diameter	.04841	.006758	**.3129****

Note: ** indexes $p<.01$, * indexes $p<.05$

Table 2 shows the regression table from this model. Interestingly, all physiological variables that were individually correlated with reaction time remained significant predictors of reaction time in the multiple regression model. Note also that in additional

analyses, entering the variables in different orders into the regression model in all cases somewhat changed the beta weights, but all variables remained significant predictors even when entered into the model last (i.e., when variance due to all other variables had been accounted for).

The hypothesized dissociation, specifically between the relationships of pupil size latency and P300 latency to reaction time, was therefore not supported by our statistical analysis. Rather, both P300 latency and pupil dilation latency continued to predict reaction time when the respective other variable had been accounted for.

Correlations Between Physiological Measures and Reaction Time During the Recognition Test. Table 3 displays the partial correlation of each measure with log reaction time at test, when variance due to *participant* and *stimulus type* has been partialed out. Significant negative correlations with reaction time at test were found for Novelty P3 amplitude and P300 amplitude (for the P300 factor from the first PCA); that is, larger amplitudes at encoding were associated with shorter reaction times at test. In addition, mean pupil size in the return-to-baseline time window was positively correlated with reaction time at test: Slower returns to baseline were associated with slower responses at test.

The Novelty P3 and P300 patterns are consistent with the statistical trends we found for the median split analysis of reaction time at recognition (although these were previously non-significant). Furthermore, the correlation for the return of the pupil diameter to baseline to recognition speed was not significant in the previous analysis, although there had been a trend for an interaction between stimulus type and subsequent recognition speed for this measure.

Table 3. Correlations among log reaction time at test and physiological variables elicited at encoding, when variance due to *participant* and *stimulus type* has been partialed out.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Log RT (Recogn.)	-										
2. Novelty P3 Max	**-.1**** ***-.18***** *-.05* *-.08*	-									
3. Novelty P3 Latency	-.02 *-.04* *.01* *-.04*	**.1**** ***.16***** *.07* *.11*	-								
4. P300 Max (1st PCA)	**-.08*** ***-.16***** *-.06* *.03*	**.37**** ***.48***** ***.41***** ***.24*****	.04 *.05* *.08* *.03*	-							
5. P300 Latency (1st PCA)	.03 *.02* *.05* *-.01*	-.04 *-.00* *-.08* *.07*	**.16**** ***.17***** ***.19***** *.08*	.04 *.02* *.01* *.14*	-						
6. P300 Max (2nd PCA)	-.04 *.02* *-.07* *-.12*	**.4**** ***.37***** ***.39***** ***.4*****	**.08*** *.03* *.08* *.1*	**.38**** ***.42***** ***.46***** ***.24*****	.04 *-.01* *.03* *.13*	-					
7. P300 Latency (2nd PCA)	-.04 *.03* *-.06* *-.08*	-.02 *.04* *-.05* *.01*	**.16**** *.11* ***.16***** *.14*	-.03 *.03* *-.03* *-.07*	**.3**** ***.3***** ***.37***** *.05*	.01 *.06* *-.02* *.03*	-				
8. Pupil Dilation Max	-.01 *-.03* *-.02* *.02*	-.01 *.01* *.05* *.01*	-.01 *-.01* *.07* *-.11*	-.00 *.00* *-.03* *-.05*	.02 *.01* *.06* *-.14*	-.03 *-.09* *.02* *-.01*	.04 *-.01* *.06* *.02*	-			
9. Pupil Dilation Latency	.02 *.01* *.01* *.07*	-.04 *-.05* *-.07* *.07*	.05 *.03* *.07* *-.05*	-.06 *-.04* ***-.1**** *-.06*	**.08*** ***.16***** *.09* *-.1*	-.03 *-.06* *-.05* *-.03*	.04 ***.13***** *.00* *-.1*	**.12**** ***.14**** ***.11**** ***.15*****	-		
10. Pupil Mean 2nd TW	**.07*** *.04* *.08* *.02*	-.03 *.00* *-.02* *-.04*	-.00 *.01* *-.02* *-.13*	-.02 *.02* *-.07* *-.04*	.03 *.02* *.06* *-.05*	-.06 *-.1* *-.08* *-.00*	.04 *.04* *.03* *.04*	**.8**** ***.81***** ***.8***** ***.81*****	**.31**** ***.35***** ***.3***** ***.28*****	-	
11. Pupil Baseline	.04 *.06* *.07* *.01*	-.06 *-.09* ***-.15***** ***.01***	-.00 *-.07* *-.00* *.06*	**-.08**** *-.04* ***-.15***** *.03*	.03 *.00* *.06* *.03*	-.02 *-.06* *.00* *-.04*	-.00 *-.01* *.03* *-.08*	**-.43**** ***-.41***** ***-.43***** ***-.45*****	-.02 *-.02* *-.03* *.01*	**-.43**** ***-.41***** ***-.42***** ***-.48*****	-

Only trials are included that were associated with subsequent confident hits. The first number in each cell is the partial correlation coefficient between the two measures with *stimulus type* partialed out. The remaining three numbers represent the correlation coefficient for (1) frequents, (2) infrequents or (3) pictures only. Shaded cells in the first column index physiological measures correlated (across stimulus types) with log reaction time at test. Shaded areas in the rest of the table indicate correlation coefficients between two variables that are both individually correlated with reaction time at test. *Note:* ** indexes $p<.01$, * indexes $p<.05$

Table 4. Multiple regression model on the log reaction times at test, which included *participant* as a random covariate, as well as *stimulus type* as fixed covariate.

Variable	Unstandardized Coefficient (B)	Standard Error	Standardized Coefficient (β)
Model 1			
Novelty P3			
Amplitude	-.00168	.000762	**-.08144***
P300 (1st PCA)			
Amplitude	-.00067	.000555	-.04761
Pupil			
Mean dilation 1.5-2s	.04084	.02077	**.06084***
Model 2			
P300 (1st PCA)			
Amplitude	-.00112	.000517	**-.07936***
Pupil			
Mean dilation 1.5-2s	.04216	.02081	**.06280***
Model 3			
Novelty P3			
Amplitude	-.00202	.000709	**-.09798****
Pupil			
Mean dilation 1.5-2s	.04091	.02077	**.06094***

Note: ** indexes $p<.01$, * indexes $p<.05$

We entered the three physiological variables that were individually correlated with reaction time at test into a multiple regression model along with the covariates *participant* and *trial type*. As shown in table 4 (model 1), Novelty P3 amplitude and pupil diameter in the second time window remained significant predictors of reaction time at test, whereas P300 amplitude was non-significant. To investigate whether the P300 was not a significant predictor in this model due to the shared variance between Novelty P3 and P300 amplitude (the correlation between the two variables was $r=.34$, table 3), we ran two additional analyses that included either Novelty P3 or P300 amplitude as a predictor. As shown in table 4 (models 2 and 3), larger P300- and Novelty P3 amplitudes at encoding predicted shorter reaction times when the respective other ERP component was not included in the model. Thus, both Novelty P3- and P300 amplitude predicted

subsequent recognition reaction time when pupil diameter in the "return to baseline" time window had been accounted for, but while the Novelty P3 continued to predict reaction time at test when P300 amplitude had been accounted for, this was not true vice versa.

One complication when interpreting these results is that some stimuli might be inherently easier to process than others, leading to variance in reaction times at recognition that is not directly related to memory strength. Processing difficulty is also likely to affect the physiological measures, which could lead to a correlation between physiological measures and subsequent recognition reaction time that does not, as would be of interest, index associations with subsequent memory strength. To test this idea, we assumed that variance due to inherent stimulus characteristics would influence reaction time both at encoding and at test. To account for this, we modified our models to include log reaction time at encoding as an additional predictor of reaction time at recognition.

Table 5 displays the main results of this analysis. As expected, reaction time at encoding was highly predictive of reaction time at test in all of the models. Importantly, after including reaction time at encoding as a predictor, mean pupil diameter in the "return-to-baseline" time window no longer significantly predicted reaction time at test in any of our models. However, both Novelty P3 amplitude and P300 amplitude, at least when included in separate regression models (models 2 and 3, table 5) continued to predict subsequent reaction times at test.

This finding was also corroborated by an examination the partial correlations of each physiological variable with reaction time at test when *participant*, *stimulus type* and *reaction time at encoding* had been partialed out. Thus, the partial correlations with reaction time at test remained significant for Novelty P3 amplitude (r=-.08, p=.01) and

P300 amplitude measured in the first PCA (r=-.07, p=.03), but not for the mean pupil diameter in the second time window (r=.05, p>.1). It is also worth noting that when variance due to the reaction time at encoding was partialed out, still none of the other physiological measures showed significant partial correlations with recognition speed.

Table 5. Multiple regression model on the log reaction times at test, which included *participant, stimulus type* and *reaction time at encoding* as covariates.

Variable	Unstandardized Coefficient (B)	Standard Error	Standardized Coefficient (β)
Model 1			
RT at Encoding	.1637	.04908	**.1198****
Novelty P3			
Amplitude	-.00129	.000766	-.06245
P300 (1st PCA)			
Amplitude	-.00072	.000552	-.05116
Pupil			
Mean dilation 1.5-2s	.03271	.02080	.04872
Model 2			
RT at Encoding	.1762	.04857	**.1289****
P300 (1st PCA)			
Amplitude	-.00106	.000514	**-.07516***
Pupil			
Mean dilation 1.5-2s	.03308	.02083	.04928
Model 3			
RT at Encoding	.1620	.04908	**.1185****
Novelty P3			
Amplitude	-.00166	.000713	**-.08035***
Pupil			
Mean dilation 1.5-2s	.03287	.02081	.04869

Note: ** indexes p<.01, * indexes p<.05

Overall, our results suggest that only Novelty P3 amplitude and P300 amplitude in the spatial factor obtained from the first PCA are predictive of subsequent memory strength, as indexed by reaction time during recognition. However, their amplitudes appear to account for overlapping portions of the variance in recognition reaction times.

Table 6. Correlations among log reaction time to the *next* trial and physiological variables elicited at encoding, when variance due to *participant, task, sub-block* and *stimulus type* has been partialed out.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Log RT (Next Trial)	-										
2. Novelty P3 Max	.01 *.03* *-.06* *.04*	-									
3. Novelty P3 Latency	-.04 ***-.05**** *.06* *-.04*	**.1**** ***.1**** *.04* *.1*	-								
4. P300 Max (1st PCA)	.03 *.03* *.00* *.01*	**.39**** ***.39**** ***.44**** ***.32****	.04 *.04* *.1* *-.01*	-							
5. P300 Latency (1st PCA)	.04 *.04* *.01* *-.04*	-.00 *-.01* *-.07* *.08*	**.11**** ***.09**** ***.18**** *.05*	**.05*** *.05* *.01* ***.16****	-						
6. P300 Max (2nd PCA)	.04 *.04* *.06* *.00*	**.34**** ***.32**** ***.4**** ***.46****	.04 *.02* *.07* *.14*	**.46**** ***.49**** ***.46**** ***.32****	.03 *.03* *-.01* ***.19****	-					
7. P300 Latency (2nd PCA)	.03 *.02* *-.02* *-.04*	-.02 *-.01* *-.09* *.03*	**.13**** ***.12**** ***.17**** *.1*	.01 *.05* *-.03* *-.12*	**.35**** ***.36**** ***.39**** *.00*	.03 *.05* *.01* *-.12*	-				
8. Pupil Dilation Max	.01 *.02* *-.01* *.01*	-.03 *-.05* *.07* *-.02*	-.01 *-.01* *.02* *-.04*	-.03 *-.02* *-.03* *-.01*	.03 ***.05**** *-.01* ***-.16****	-.03 *-.02* *-.05* *-.03*	**.05*** ***.05**** *.01* *.05*	-			
9. Pupil Dilation Latency	**.06**** ***.08**** *-.08* ***.24****	**-.04*** *-.04* *-.08* *.09*	.01 *-.01* *.05* *-.04*	-.03 *-.03* *-.09* *-.02*	**.04*** *.05* *.06* ***-.19****	-.04 *-.05* *-.07* *.04*	.01 *.02* *-.00* *-.04*	**.11**** ***.11**** ***.13**** *.11*	-		
10. Pupil Mean 2nd TW	.02 *.03* *-.04* *.1*	**-.05*** *-.04* *.00* *-.07*	-.01 *-.01* *.01* *-.02*	**-.05*** *-.03* *-.08* *.01*	.03 *.03* *.02* *-.11*	**-.06**** ***-.06**** ***-.14**** *.01*	.03 *.04* *.02* *.08*	**.78**** ***.78**** ***.8**** ***.8****	**.3**** ***.29**** ***.29**** ***.26****	-	
11. Pupil Baseline	.04 ***.05**** *.00* *.02*	.00 *.02* ***-.12**** *.06*	-.04 ***-.07**** *.03* *.04*	-.03 *.00* *-.11* *-.06*	.03 *.00* ***.12**** *.05*	.02 *.02* *.04* *-.02*	-.01 *-.03* *.09* *-.08*	**-.39**** ***-.37**** ***-.43**** ***-.47****	.00 *.02* *-.08* *-.02*	**-.42**** ***-.41**** ***-.43**** ***-.48****	-

Only trials are included that were followed by a frequent trial. The first number in each cell is the partial correlation coefficient between the two measures with *stimulus type* partialed out. The remaining three numbers represent the correlation coefficient for (1) frequents, (2) infrequents or (3) pictures only. Shaded cells in the first column index physiological measures correlated (across stimulus types) with log reaction time at test. Shaded areas in the rest of the table indicate correlation coefficients between two variables that are both individually correlated with reaction time at test. *Note:* ** indexes $p<.01$, * indexes $p<.05$

Correlations between Physiological Measures and Reaction Time on the Next Trial. The final question concerned the relationship between physiological measures and the reaction time to the next trial (if the next trial was of the frequent category). As shown in table 6, the only significant correlation was between pupil dilation latency and reaction time to the next trial: The later the peak dilation occurred, the longer was the reaction time on the next trial. Since no other correlations were significant, it was unnecessary to run a regression analysis analogous to those reported in the previous sections.

Additional Findings from the Correlation/Regression Analysis. There are several additional patterns within the correlational structure of our data that are worth noting and discussing in the context of our research questions and prior literature. Of interest are correlations between the different physiological measures (tables 1, 3 and 6). Note that because a different subset of the trials is included in each table (for example, table 6 only includes trials that were followed by frequents), the precise values are slightly different, but the general patterns are very similar across tables. Since calculations of the correlations in table 1 included the largest number of trials we will refer to table 1 here.

First, ERP amplitudes of Novelty P3, P300 from the first PCA and P300 from the second PCA were relatively highly correlated with each other (the smallest correlation was between Novelty P3 and P300 from the second PCA: r=.34). However, the pupil dilation amplitude measured in the first time window was generally *uncorrelated* with ERP amplitudes from either PCA factor. Small, but significant *negative* correlations were found between the mean diameter in the second time window and ERP amplitudes (r ranged between -.04 and -.06). If the pupil dilation response reflected the same cognitive

process as any of the ERP components, a strong positive correlation would be expected. Our correlational patterns speak against this idea.

Also worth discussing is the relationship between baseline pupil diameter on the one hand, and reaction time, pupil dilation magnitude and P300 amplitude on the other hand. Thus, the adaptive gain theory of the PDR (e.g., Gilzenrat et al., 2010) and the P300 (e.g., Nieuwenhuis, Aston-Jones, et al., 2005) predicts an association of smaller baseline diameters with (1) better performance, as for example indexed by shorter reaction times, (2) larger pupil dilation amplitudes and (3) larger P300 amplitudes. Table 1 clearly support points (1) and (2), but the support for point (3) within our dataset is weak. That is, baseline diameter was overall uncorrelated with any ERP amplitude measure (r ranged between 0 and -.03). Only when just infrequent trials were considered, a significant correlation in the predicted direction was found between baseline diameter on the one hand, and Novelty P3 amplitude (r=-.1) and P300 amplitude from the first PCA (r=-.11) on the other.

Discussion

The main patterns in the behavioral data were that (1) words of the infrequent category were associated with longer reaction times and higher error rates than words in the frequent category and images that depicted objects of the frequent category, and (2) pictures were associated with the strongest memory traces, followed by infrequents and finally frequents, as indexed by recall rates, recognition sensitivity, as well as reaction times to correct, high confidence "old" (hits) and "new" (correct rejections) responses.

The behavioral patterns at encoding could have two different explanations, which, based on our experimental design, cannot be easily disentangled. In our view, the more likely explanation is that words of the infrequent category were the only stimuli that required an "infrequent response", in other words a response switch. Thus, since in 86% of the trials, one response was required, it was useful for the participants to pre-program this frequent response. If, in contrary to a developed expectation, an infrequent stimulus was presented, this pre-programmed response needed to be inhibited and the other response needed to be prepared. This response switch would have led to the observed increase in reaction time and error rate.

The alternative explanation is that stimuli of the frequent category (including the pictures) benefitted from a semantic priming effect. Thus, the previous presentation of a large number of stimuli of the same category might have facilitated processing of frequents and pictures, while infrequents were not preceded by as many stimuli from the same category and therefore did not benefit from such a priming effect. Our experimental

design confounds semantic deviancy with response switching, so we cannot completely rule out this possibility. However, if priming took place in our design, we should observe a larger amplitude in the ERP component known as the N400 (Kutas & Hillyard, 1980) for those stimuli that did not benefit from the priming effect, that is, the infrequents. However, although negative peaks at 400ms after the stimulus were observed in several components that showed centro-parietal distribution consistent with an N400 distribution (for example, figure 7 A, SFs 2, 4 and 5; see also the grand averages in figure 5), none exhibited a larger negative peak at 400ms for infrequent stimuli. The lack of N400 (or priming) differences between stimulus types might be due to the fact that our categories were very broad. We therefore consider the response switching explanation more likely than the priming explanation.

The memory patterns including recall rates, recognition accuracy and reaction time during the recognition test, suggest that pictorial stimuli were encoded into the strongest memories, and that participants might have used retrieval strategies that favored the pictures. This pattern is in line with a large body of literature on the "picture superiority effect" (e.g., Shepard, 1967). There is some evidence that the picture superiority effect occurs due to an enhanced encoding of distinctive details of pictorial vs. verbal stimuli (Curran & Doyle, 2011). Furthermore, our recall and recognition memory data also suggest that infrequents were more likely to be encoded and retrieved both in the recall and recognition tests, compared to frequents. Again, the memory superiority of infrequent stimuli is in line with a large body of previous literature (e.g., Hunt, 2006; Von Restorff, 1933).

It is important to take into account these behavioral patterns when interpreting the differences in the physiological responses between stimulus types and their correlation with behavioral data (such as reaction time at encoding and at test). Thus, combining patterns within the physiological responses with the behavioral differences between stimulus types can help constrain the interpretations of the physiological responses' functional significance.

P300, Reaction Time and Subsequent Memory

Before discussing the relationship between P300 and behavioral patterns, it first necessary to discuss the fact that we obtained two PCA factors that showed typical characteristics of a P300 – a parietal/posterior spatial distribution as well as a large positive peak for infrequents in the time window of 500 to 900ms. It is possible that the two factors reflect two different scalp-recorded ERP components, possibly with distinct neural correlates and functions. However, an alternative possibility, which in our view is more likely, is that the two P300-like factors are a by-product of the analysis method used. PCA aims to explain as much variance as possible with a small number of factors. However, the rotation method attempts to minimize loadings of intermediate magnitude and maximize larger loadings, effectively "focusing" the highest spatial factor loadings onto a relatively small scalp area (this is true for both Varimax and Promax, for an explanation of the rotation methods, see Dien, 2010b). Especially when a scalp-recorded ERP component has a broad scalp distribution, it is therefore possible that the variance induced by a single ERP component is captured by two PCA factors. After initially obtaining the two P300-like factors from the PCA, we did not take a strong stance as to whether the two factors represent the same- or different ERP components. Rather, we

included both factors in our analysis. If the factors had varied in qualitatively different ways with our experimental manipulations and behavioral data, this would have spoken against the idea that both factors capture the same ERP component. However, in almost all of our analyses, the two factors showed similar response patterns: Both factors showed a large, early positivity for pictures and a broader positivity between 500 and 900ms post-stimulus that was larger for infrequents than frequents. Similarly, the latencies of both factors were correlated with reaction time, and both factors' amplitudes were correlated with subsequent recall. Furthermore, the amplitudes (r=.46) and latencies (r=.34) of both factors were also relatively highly correlated with each other on an individual trial basis (note that on individual trials of EEG activity there are high levels of noise, so these correlation coefficients can be considered relatively high). In the following, we will therefore focus on the discussion of the factor from the first PCA on all stimulus types, as this factor explained a larger portion of the variance in the data.

The relatively late P300 peak for infrequents (at about 700ms after the stimulus) is not untypical, as we used verbal stimuli – that is, it is reasonable to assume that it takes a longer time to detect semantic deviance as opposed to deviance based on more superficial stimulus characteristics. However, the positivity that pictures elicited in the P300 factor was much sharper and peaked much earlier than the P300 for the infrequents. In principle, latency differences between deviance-related components elicited by infrequents and pictures are not surprising since it should take longer to register semantic deviance than perceptual deviance. However, a P300 that peaks about 200ms after stimulus onset would be extremely early. Furthermore, it is inconsistent with prior studies (Spencer et al., 1999) and therefore unlikely that such strong latency differences would

be observed in the P300 but not in the Novelty P3. In summary, it is unclear whether in our paradigm pictures elicited a P300.

The finding that P300 latency was consistently correlated with reaction time at encoding is a replication of a large number of prior findings, beginning in the study of Kutas et al. (1977) where response accuracy was encouraged. However, disconfirming our hypothesis, P300 latency continued to statistically predict reaction time on individual trials when all other physiological responses had been accounted for. In other words, statistically, P300 latency explained a different portion of the variance in reaction time than, for example, the pupil dilation response.

In the light of the prior findings reviewed in the introduction, which showed that P300 latency and reaction time could be dissociated, it is important to note that our finding should not be taken as evidence for a direct relationship between P300 and behavioral responding. All that can be concluded from our results is that P300 latency and PDR latency do not explain the same portion of the reaction time variance. One hypothesis that should be tested in further studies is that P300 latency explains variance in reaction times that is due to the stimulus evaluation processes, whereas the PDR explains variance due to aspects of response preparation- or execution. This idea could be better tested in an experimental design that manipulates stimulus evaluation- and response demands independently.

The P300 subsequent memory effect – larger P300 amplitudes for subsequently recalled- compared to not recalled stimuli when elaborative encoding is not used – also replicated many prior studies, starting with Karis et al. (1984). However, the present study is the first to also demonstrate, on an individual trial basis, a correlation between

P300 amplitude during encoding and *reaction time* during a recognition test. The correlation remained significant when variance due to reaction time at encoding, which we used as an index of inherent differences in processing time between stimuli, was accounted for. Taken together, the traditional P300 subsequent memory analysis and the correlation with recognition reaction time converged on the idea that P300 amplitude elicited at encoding correlates with subsequent memory. One possible interpretation is that P300 amplitude is proportional to subsequent memory strength, which should facilitate both recall and recognition.

The Novelty P3 as an Index of Resource Allocation?

The centrally distributed factor was interpreted as a Novelty P3, based on its spatial distribution and the classical Novelty P3-like response to experimental manipulation: the positivity was largest for perceptually deviant, infrequent stimuli, second largest for semantically deviant, but not perceptually salient words, and smallest for words of the frequent category. While the stimuli themselves were not task-irrelevant as in the typical Novelty P3 oddball paradigm (Courchesne et al., 1975, but see Cycowicz & Friedman, 2004), the presentation format (i.e., the fact the respective noun was presented in picture form) *was* task-irrelevant. Our findings are thus in line with the idea that a Novelty P3 can be elicited even when a response is required to the eliciting stimulus, which speaks against the response inhibition hypothesis of the Novelty P3. It is therefore likely that the No-Go P3 (e.g., Pfefferbaum et al., 1985) is a functionally separate component to the Novelty P3 observed in our paradigm.

The pattern in Novelty P3 amplitude did not qualitatively parallel the behavioral measures from the encoding phase: reaction times were slowest and error rates highest

for infrequents, but these showed an intermediate Novelty P3 amplitude. Performance levels (reaction time and error rate) were about equal for frequents and pictures, yet the Novelty P3 amplitudes elicited by these two stimulus types vastly diverged. Overall, in light of our data it seems unlikely that the Novelty P3 elicited in our paradigm indexes response-related processes.

Novelty P3 amplitude was correlated with reaction time at encoding in both the median split- and the individual trial analysis. Interestingly, in its variance with stimulus type, Novelty P3 amplitude also directly mirrored subsequent memory strength. That is, pictures both elicited the largest Novelty P3 and were associated with the strongest memory traces. Infrequents elicited the second largest amplitude and exhibited the second strongest memory traces. Finally, frequents elicited the smallest amplitudes and were associated with the weakest memories (at least to the extent to which strength is indexed by recall probability, recognition sensitivity and/or reaction time during the recognition test). Furthermore, within stimulus types, larger Novelty P3 amplitudes were associated with larger probabilities of subsequent recall and, in individual trials, shorter reaction times during the subsequent recognition test. Taking all these patterns together, one possible interpretation is that Novelty P3 amplitude might reflect resource allocation to a given trial. That is, the more resources are allocated to an experimental trial (1) the quicker the response (at least within stimulus types) and (2) the stronger this trial is encoded into episodic memory.

In a dual-task paradigm, Wickens and colleagues (1983) varied the difficulty of the 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 extent

to which resources were allocated to the oddball task. Since in such early studies dense electrode arrays were not available, and due to the strong spatio-temporal 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 centrally distributed Novelty P3, as obtained in the present study. This issue remains to be investigated in future studies.

In sum, we found no support for our (very exploratory) hypothesis that Novelty P3 amplitude would be related to immediate responding. Based on our data it appears more likely that Novelty P3 amplitude is sensitive to processes elicited by deviance that operate in parallel to the stimulus-response stream, such as resource allocation.

The Pupil Dilation Response and Behavioral Responding

The current view in the literature is that a temporary increase in pupil size – the pupil dilation response – is elicited by events that deviate from expectancies, such as infrequent events in an oddball paradigm (Murphy et al., 2011), behavioral errors (Wessel et al., 2011), or the delivery of an unexpected reward or the absence of an expected reward (Preuschoff et al., 2011). The present study suggests that the PDR is not elicited by low stimulus probability or deviancy due to perceptual characteristics *per se* – if this was the case, pictures should have elicited larger amplitudes than frequents (recall that pictures were equally improbable as infrequents). In contrast, infrequents elicited the largest PDR, in our view most likely due to the associated response switch.

The idea that response-related processes are reflected in the PDR was further supported by the correlation of its latency and its mean amplitude in the "return to baseline" time window with reaction time – longer PDR latencies and larger amplitudes

in the second time window were associated with longer reaction times on the same trial. Note that this finding is in contrast to Gilzenrat (2010), who found larger PDR amplitudes to be correlated with *better* performance. However, it is principally in line with Murphy et al. (2011) in that larger amplitudes (at least in the "return to baseline" time window) were associated with longer reaction times.

Murphy et al. (2011) also found that trials in which PDRs were large (and performance was relatively low) were *followed* by improvements in performance. In our single trial correlational analysis, there was some evidence for a slower response to the next trial when PDR *latency* was relatively long. However, we did not find any correlations between PDR *amplitude* and reaction time to the next trial. The reasons for this discrepancy between our studies and others are unclear, but it is important to point out one major difference between our oddball task and Murphy et al's and Gilzenrat et al's: In both of their studies, the oddball tasks used very simple and easy to categorize stimuli, while our study involved a relatively complex semantic judgment.

Overall, our data therefore support the view that the cognitive process indexed by the PDR is related to response preparation or execution, in line with Nieuwenhuis et al. (2011). However, our data do not support the idea that larger PDR amplitudes directly correspond to better task performance (see, for example Gilzenrat et al. 2010).

Nieuwenhuis (2011) recently proposed that phasic LC activity and consequently the PDR (along with the P300) are not only related to behavioral responding, but also to memory encoding. Our analyses provided only weak evidence for a correlation between pupil measures and subsequent memory strength. There was a small subsequent memory effect, such that larger PDR amplitudes were associated with a higher probability of

subsequent recall. However, this finding did not appear to be statistically robust, as it did not hold up for analyses conducted separately for each stimulus type. In the individual trial analysis, we did obtain a correlation between the mean amplitude in the "return to baseline" time window and reaction time at test. However, this correlation became non-significant when reaction time at encoding to the same stimuli – which can index response speed independently of memory strength – was included as a predictor. It seems therefore, that pupil diameter measures co-vary with the extent to which the eliciting stimulus will later be remembered, but that this correlation is not due to a direct link between the cognitive process indexed by the PDR and memory encoding. The associations might be explained by third variables (possibly stimulus-evaluation or response-related processes) that affect both the PDR and subsequent memory strength.

Relationship between P300, Novelty P3 and PDR

In this section we will review the empirical evidence obtained in our study in the light of the adaptive gain theory of LC function (especially in its implications for PDR and P300) as well as the context updating hypothesis of the P300. First, it is important to re-iterate that PDR amplitude was uncorrelated with P300 or Novelty P3 amplitude. In fact, pupil size in the "return to baseline" time window even *negatively* correlated with ERP amplitudes. Furthermore, the response patterns for frequents, infrequents and pictures were not exactly in parallel between the PDR and the ERP components (keeping in mind, however, that it is not clear whether the positivity elicited by pictures in the P300 factor is an instance of the P300). These patterns suggest that while P300, Novelty P3 and the PDR are elicited by events that violate expectancies, the cognitive processes indexed by

each response are unlikely to be "central and autonomic analogues of the same cognitive process" (Nieuwenhuis et al., 2011).

The extent to which our data confirm predictions from the adaptive gain theory (Aston-Jones & Cohen, 2005) can also be investigated by exploring relationships between the tonic pupil diameter (as approximated by pupil diameter in the baseline period) on the one hand, and measures of performance, PDR amplitude and P300 amplitude on the other. As predicted by the adaptive gain theory, on individual trials smaller baseline diameters were associated with faster reaction times – especially for frequent- and picture trials – as well as larger PDR amplitudes in the first time window and in the "return to baseline" time window. While in principle, this pattern could be explained such that larger baseline diameters do not allow for as much dilation as smaller baseline diameters because of ceiling effects, Gilzenrat et al. (2010) found the same pattern and through elegant supplemental analyses ruled out the ceiling effect possibility.

In the light of these significant and relatively high (taking into account the relatively high levels of noise on individual trials) correlations, it is even more striking that baseline diameter was uncorrelated with P300 amplitude. This speaks against the idea that the P300 reflects the phasic response in the LC, at least to the extent to which P300 is assumed to follow the patterns predicted by the adaptive gain theory. It is worth noting, however, that Novelty P3 amplitude *was* significantly and negatively correlated with baseline pupil diameter (r=-.04, table 1). It is possible that prior studies that did not use PCA to disentangle ERP components, and that reported negative correlations between baseline pupil diameter and P300 amplitude (e.g., Murphy et al. 2011), in fact picked up variance from the Novelty P3. The other patterns within the Novelty P3 were also

consistent with predictions from the adaptive gain theory: Its response patterns paralleled memory strength, its amplitude was negatively correlated with reaction time on the same trial and at test, and subsequent memory effects were present for both subsequent recall and subsequent recognition speed. All these associations between Novelty P3 and behavior are generally in line with the idea that large Novelty P3 amplitudes coincide with a strong task focus ("exploitation"), which, in turn leads to improved task performance and memory encoding.

It is next worth discussing separately each behavioral measure that we investigated. For example, many physiological variables were correlated with reaction time on the same trial. In the individual trial analysis, these included Novelty P3 amplitude, P300 latency, PDR latency, mean pupil diameter in the "return to baseline" time window and baseline pupil diameter. All variables remained significant predictors of reaction time when the other physiological responses had been accounted for. Therefore, our data do not allow for the conclusion that the PDR is more closely related to reaction time than P300, as we had hypothesized. However, it is also not warranted to conclude that all of our physiological responses are indices of cognitive processes that are directly integrated into the stimulus-response stream. For example, the link between P300 latency and reaction time has been found previously (Kutas et al., 1977) when, like in our study, participants were instructed to respond accurately (the high accuracy of .94 confirms that participants in our study attempted to respond accurately). However, in a condition where participants were asked to respond faster, the correlation was abolished (Kutas et al. 1977). To clarify which response is more closely associated with behavioral responding, future studies should apply experimental paradigms that differentially manipulate

variance in reaction time due to stimulus evaluation time and due to response demands. If the former manipulation primarily affected P300 latency and the latter primarily affected PDR latency, this would speak for a dissociation of the extent to which each response is associated with behavioral responding.

It is also worth noting that the *amplitudes* of neither P300 nor PDR were negatively correlated with reaction time. This constrains the theory of their functional significance such that the extent to which the associated cognitive processes are elicited does not appear to affect behavior directly.

The context updating hypothesis predicts that P300 amplitude will be correlated with the probability of subsequent recall for the eliciting event (Donchin, 1981). In line with this idea and with many prior studies beginning with Karis et al. (1984), we obtained a P300 subsequent memory effect. The correlation between P300 amplitude and the reaction time during recognition strengthened the association between P300 and memory. Interestingly, however, Novelty P3 amplitude was also correlated with subsequent recall (in line with Kamp et al., in press) and recognition speed, and the regression analysis suggested that the two ERP components accounted for overlapping portions of the variance in subsequent memory.

The evidence for a correlation between PDR and subsequent memory was much weaker, as discussed above. Unlike the PDR, the P300 and Novelty P3 remained significant predictors of subsequent recognition reaction time when reaction time at encoding had been accounted for, suggesting a stronger (and possibly more direct) relationship between the ERP components and subsequent memory than for the PDR.

Summary and Conclusions

Overall, the results of the present study suggest that the Novelty P3, P300 and the PDR respond in different ways to experimental manipulations and co-vary in qualitatively different ways with behavior. None of our physiological measures uniquely predicted reaction time or subsequent memory, but our data speak for a closer correlation of P300 with subsequent memory and a closer correlation of the PDR with behavioral responding. Novelty P3, in contrast, co-varied with many measures of behavior, leading us to suggest that it might index a more general resource allocation process.

To further investigate these relationships, future studies should employ paradigms that manipulate stimulus evaluation time and response preparation demands independently. Such designs may be more sensitive to disentangle the variance of each measure with the two processes.

List of References

Arbel, Y., & Donchin, E. (2009). Parsing the componential structure of post-error ERPs: A principal component analysis of ERPs following errors. *Psychophysiology, 46*(6), 1179-1189. doi: 10.1111/j.1469-8986.2009.00857.x

Arbel, Y., Spencer, K., & Donchin, E. (2010). The N400 and the P300 are not all that independent. *Psychophysiology, 48*(6), 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. *Annu. Rev. Neurosci., 28*, 403-450.

Axmacher, N., Cohen, M.X., Fell, J., Haupt, S., Duempelmann, M., Elger, C.E., . . . Ranganath, C. (2010). Intracranial EEG correlates of expectancy and memory formation in the human hippocampus and nucleus accumbens. *Neuron, 65*(4), 541-549. doi: 10.1016/j.neuron.2010.02.006

Azizian, A., & Polich, J. (2007). Evidence for attentional gradient in the serial position memory curve from event-related potentials. *J Cogn Neurosci*, *19*(12) 2071-2081. doi: 10.1162/jocn.2007.19.12.2071

Barry, R.J. (2009). Habituation of the orienting reflex and the development of preliminary process theory. *Neurobiology of Learning and Memory, 92*(2), 235-242. doi: 10.1016/j.nlm.2008.07.007

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*(10), 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*(3), 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*(2), 131-143. doi: 10.1016/0013-4694(75)90003-6

Curran, T., & Doyle, J. (2011). Picture superiority doubly dissociates the ERP correlates of recollection and familiarity. *Journal of Cognitive Neuroscience, 23*(5), 1247-1262.

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*(1), 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 : official journal of the International Federation of Clinical Neurophysiology, 115*(6), 1359-1367.

Cycowicz, Y.M., & Friedman, D. (2007). Visual novel stimuli in an ERP novelty oddball paradigm: Effects of familiarity on repetition and recognition memory. *Psychophysiology, 44*(1), 11-29. doi: 10.1111/j.1469-8986.2006.00481.x

Dien, J. (2010a). The ERP PCA toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods, 187*(1), 138-145.

Dien, J. (2010b). Evaluating two−step pca of ERP data with geomin, infomax, oblimin, promax, and varimax rotations. *Psychophysiology, 47*(1), 170-183.

Dien, J., Spencer, K.M., & Donchin, E. (2003). Localization of the event-related potential novelty response as defined by principal component analysis. *Cognitive Brain Research, 17*(3), 637-650.

Donchin, E. (1981). Surprise! ... Surprise? *Psychophysiology, 18*(5), 493-513.

Donchin, E., & Coles, M.G.H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 11*(03), 357-374. doi: 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*(1), 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: Tilburg University Press.

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., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. *Biological Psychology, 14*(1–2), 1-52. doi: http://dx.doi.org/10.1016/0301-0511(82)90016-3

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. *J Exp Psychol Learn Mem Cogn, 21*(1), 224 - 240.

Fabiani, M., & Friedman, D. (1995). Changes in brain activity patterns in aging: The novelty oddball. *Psychophysiology, 32*(6), 579-594. doi: 10.1111/j.1469-8986.1995.tb01234.x

Fabiani, M., Karis, D., & Donchin, E. (1986). P300 and recall in an incidential memory paradigm. *Psychophysiology, 23*(3), 298-308.

Fabiani, M., Karis, D., & Donchin, E. (1990). Effects of mnemonic strategy manipulation in a von Restorff paradigm. *Electroencephalogr Clin Neurophysiol, 75*(2), 22 - 35.

Fernández, G., Effern, A., Grunwald, T., Pezer, N., Lehnertz, K., Dümpelmann, M., . . . Elger, C.E. (1999). Real-time tracking of memory formation in the human rhinal cortex and hippocampus. *Science, 285*(5433), 1582-1585. doi: 10.1126/science.285.5433.1582

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

Friedman, D. (1984). P300 and slow wave: The effects of reaction time quartile. *Biological Psychology, 18*(1), 49-71. doi: 10.1016/0301-0511(84)90028-0

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*(4), 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*(5), 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*(4), 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*(2), 198-208. doi: 10.1111/1469-8986.00022

Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological science, 4*(6), 385-390. doi: 10.1111/j.1467-9280.1993.tb00586.x

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*(2), 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*(2), 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*(6), 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*(5), 933-946.

Hajcak, G., McDonald, N., & Simons, R.F. (2003). To err is autonomic: Error-related brain potentials, ans activity, and post-error compensatory behavior. *Psychophysiology, 40*(6), 895-903. doi: 10.1111/1469-8986.00107

Heaver, B., & Hutton, S.B. (2011). Keeping an eye on the truth? Pupil size changes associated with recognition memory. *Memory, 19*(4), 398-405. doi: 10.1080/09658211.2011.575788

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. doi: 10.1016/j.ijpsycho.2005.10.015

Hunt, R.R. (2006). Distinctiveness and memory. In R. R. Hunt & J. B. Worthen (Eds.), (pp. 3-25): New York: Oxford Univerity Press.

Ila, A.B., & Polich, J. (1999). P300 and response time from a manual stroop task. *Clinical neurophysiology, 110*(2), 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*(8), 1608-1619. doi: 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. *The Quarterly Journal of Experimental Psychology, 64*(10), 1971-1989.

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

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science. 154*(3756), 1583-1585.

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

Karis, D., Fabiani, M., & Donchin, E. (1984). "P300" and memory: Individual differences in the von Restorff effect. *Cognitive Psychology, 16*(2), 177 - 216.

Kim, A.S.N., Vallesi, A., Picton, T.W., & Tulving, E. (2009). Cognitive association formation in episodic memory: Evidence from event-related potentials. *Neuropsychologia, 47*, 3162-3173.

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.

Knight, R.T., & Scabini, D. (1998). Anatomic bases of event-related potentials and their relationship to novelty detection in humans. *Journal of Clinical Neurophysiology, 15*(1), 3-13.

Knight, R.T., Scabini, D., Woods, D.L., & Clayworth, C.C. (1989). Contributions of temporal-parietal junction to the human auditory P3. *Brain Research, 502*(1), 109-116. doi: 10.1016/0006-8993(89)90466-6

Kuipers, J.R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience, 5*(61), 1-5. doi: 10.3389/fnhum.2011.00061

Kutas, M., & Hillyard, S.A. (1980). Reading sensless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science, 197*(4305), 792-795.

Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2011). Pupillary stroop effects. *Cognitive Processing, 12*(1), 13-21. doi: 10.1007/s10339-010-0370-z

Leuthold, H., & Sommer, W. (1993). Stimulus presentation rate dissociates sequential effects in event-related potentials and reaction times. *Psychophysiology, 30*(5), 510-517. doi: 10.1111/j.1469-8986.1993.tb02074.x

Linden, D.E.J. (2005). The p300: Where in the brain is it produced and what does it tell us? *The Neuroscientist, 11*(6), 563-576. doi: 10.1177/1073858405280524

Magliero, A., Bashore, T.R., Coles, M.G.H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology, 21*(2), 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*(4477), 77-80. doi: 10.1126/science.7444452

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

Miltner, W.H., Braun, C.H., & Coles, M.G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. *Journal of cognitive neuroscience, 9*(6), 788-798.

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*(11), 1532-1543. doi: 10.1111/j.1469-8986.2011.01226.x

Näätänen, R. (1978). Orienting and evoked potentials. In H. D. Kimmel, E. H. van Olst & K. F. Orlebeke (Eds.), *The orienting reflex in humans* (pp. 61-75). Hillsdale, NJ: Lawrence Erlbaum Associates.

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*(2).

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* (1 ed., pp. 209-222). Boston: MIT Press.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological bulletin, 131*(4), 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*(2), 162-175. doi: 10.1111/j.1469-8986.2010.01057.x

Nieuwenhuis, S., Gilzenrat, M.S., Holmes, B.D., & Cohen, J. (2005). The role of the locus coeruleus in mediating the attentional blink: A neurocomputational theory. *Journal of Experimental Psychology: General, 134*(3), 291-307. doi: 10.1037/0096-3445.134.3.291

Notebaert, W., Houtman, F., Opstal, F.V., Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: An orienting account. *Cognition, 111*(2), 275-279. doi: 10.1016/j.cognition.2009.02.002

Nuthmann, A., & Van Der Meer, E. (2005). Time's arrow and pupillary response. *Psychophysiology, 42*(3), 306-317. doi: 10.1111/j.1469-8986.2005.00291.x

Otero, S., Weekes, B., & Hutton, S. (2011). Pupil size changes during recognition memory. *Psychophysiology, 48*, 1346-1353.

Otten, L.J., & Donchin, E. (2000). Relationship between P300 amplitude and subsequent recall for distinctive events: Dependence on type of distinctiveness attribute. *Psychophysiology, 37*, 644-661.

Overbeek, T.J.M., Nieuwenhuis, S.T., & Ridderinkhof, K.R. (2005). Dissociable components of error processing: On the functional significance of the Pe vis-a-vis the ERN/Ne. *Journal of Psychophysiology, 19*(4), 319-329. doi: 10.1027/0269-8803.19.4.319

Paivio, A., Rogers, T., & Smythe, P.C. (1968). Why are pictures easier to recall than words? *Psychonomic Science, 11*(4), 137-138.

Paller, K.A., Kutas, M., & Mayes, A. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalogr Clin Neurophysiol, 67*(4), 360 - 371.

Paller, K.A., & Wagner, A.D. (2002). Observing the transformation of experience into memory. *TRENDS in Cognitive Sciences, 6*(2), 93-102. doi: 10.1016/s1364-6613(00)01845-3

Papesh, M.H., & Goldinger, S.D. (2011). Your effort is showing! Pupil dilation reveals memory heuristics. *Constructions of remembering and metacognition*, 215-224.

Peavler, W.S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology, 11*(5), 559-566. doi: 10.1111/j.1469-8986.1974.tb01114.x

Pfefferbaum, A., Ford, J.M., Weller, B.J., & Kopell, B.S. (1985). ERPs to response production and inhibition. *Electroencephalography and clinical neurophysiology, 60*(5), 423-434.

Polich, J. (1990). Probability and inter-stimulus interval effects on the P300 from auditory stimuli. *International Journal of Psychophysiology, 10*, 163-170.

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: An integrative review. *Biol Psychol, 41*(2), 103 - 146.

Preuschoff, K., 't Hart, B.M., & Einhauser, 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

Rabbitt, P. (1969). Psychological refractory delay and response-stimulus interval duration in serial, choice-response tasks. *Acta Psychologica, 30*(0), 195-219. doi: http://dx.doi.org/10.1016/0001-6918(69)90051-1

Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nat Rev Neurosci, 4*(3), 193-202. doi: 10.1038/nrn1052

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological bulletin, 86*(3), 446.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological bulletin, 114*(3), 510.

Ratcliff, R., & Murdock, B.B. (1976). Retrieval processes in recognition memory. *Psychological Review, 83*(3), 190.

Rolke, B., Heil, M., Streb, J., & Hennighausen, E. (2001). Missed prime words within the attentional blink evoke an N400 semantic priming effect. *Psychophysiology, 38*(2), 165-174. doi: 10.1111/1469-8986.3820165

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*(3), 240-250. doi: 10.1016/j.ijpsycho.2005.05.010

Rugg, M.D., & Curran, T. (2007). Event-related potentials and recognition memory. *TRENDS in Cognitive Sciences, 11*(6), 251-257. doi: 10.1016/j.tics.2007.04.004

Rushby, J.A., & Barry, R.J. (2007). Event-related potential correlates of phasic and tonic measures of the orienting reflex. *Biological Psychology, 75*(3), 248-259. doi: 10.1016/j.biopsycho.2007.03.003

Rushby, J.A., & Barry, R.J. (2009). Single-trial event-related potentials to significant stimuli. *International Journal of Psychophysiology, 74*(2), 120-131. doi: 10.1016/j.ijpsycho.2009.08.003

Schwartze, M., Rothermich, K., Schmidt-Kassow, M., & Kotz, S.A. (2011). Temporal regularity effects on pre-attentive and attentive processing of deviance. *Biological Psychology, 87*(1), 146-151. doi: 10.1016/j.biopsycho.2011.02.021

Shepard, R.N. (1967). Recognition memory for words, sentences, and pictures. *Journal of verbal Learning and verbal Behavior, 6*(1), 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*(1), 63-76.

Simons, R.F., Graham, F.K., Miles, M.A., & Chen, X. (2001). On the relationship of P3a and the novelty-P3. *Biological Psychology, 56*(3), 207-218. doi: 10.1016/s0301-0511(01)00078-3

Smulders, F.T.Y., Kenemans, J.L., Schmidt, W.F., & Kok, A. (1999). Effects of task complexity in young and old adults: Reaction time and P300 latency are not always dissociated. *Psychophysiology, 36*(1), 118-125. doi: 10.1017/s0048577299961590

Sokolov, E.N. (1963). Higher nervous functions: The orienting reflex. *Annual review of physiology, 25*(1), 545-580.

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*(3), 409-414. doi: 10.1017/s0048577299981180

Squires, K.C., Wickens, C., Squires, N.K., & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science, 193*(4258), 1142-1146. doi: 10.1126/science.959831

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. doi: 10.1016/0013-4694(75)90263-1

Sutton, S., Braren, M., Zubin, J., & John, E.R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science, 150*(3700), 1187-1188.

Von Restorff, H. (1933). Über die wirkung von bereichsbildungen im spurenfeld. *Psychologische Forschung, 18*, 299-342.

Voss, J.L., & Paller, K.A. (2009). Remembering and knowing: Electrophysiological distinctions at encoding but not retrieval. *NeuroImage, 46*(1), 280-289. doi: 10.1016/j.neuroimage.2009.01.048

Waters, W.F., & Wright, D.C. (1979). Maintenance and habituation of the phasic orienting response to competing stimuli in selective attention. In H. D. Kimmel, E. H. van Olst & K. F. Orlebeke (Eds.), *The orienting reflex in humans*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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*(10), 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*(4615), 1080-1082. doi: 10.1126/science.6879207

Wiswede, D., Rüsseler, J., & Münte, T.F. (2007). Serial position effects in free memory recall - an ERP-study. *Biological Psychology, 75*(2), 185-193. doi: 10.1016/j.biopsycho.2007.02.002