

Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes

Marnix Naber*

Philipps-University Marburg, Neurophysics,
Marburg, Germany

Stefan Frässle*

Philipps-University Marburg, Neurophysics,
Marburg, Germany

Ueli Rutishauser

Max Planck Institute for Brain Research, Department of
Neural Systems, Frankfurt am Main, Germany
Present address: Maxine-Dunitz Neurosurgical Institute,
Department of Neurosurgery,
Cedars-Sinai Medical Center, Los Angeles, CA

Wolfgang Einhäuser

Philipps-University Marburg, Neurophysics,
Marburg, Germany
Center for Interdisciplinary Research (ZiF),
Bielefeld, Germany



Declarative memories of personal experiences are a key factor in defining oneself as an individual, which becomes particularly evident when this capability is impaired. Assessing the physiological mechanisms of human declarative memory is typically restricted to patients with specific lesions and requires invasive brain access or functional imaging. We investigated whether the pupil, an accessible physiological measure, can be utilized to probe memories for complex natural visual scenes. During memory encoding, scenes that were later remembered elicited a stronger pupil constriction compared to scenes that were later forgotten. Thus, pupil size predicts success or failure of memory formation. In contrast, novel scenes elicited stronger pupil constriction than familiar scenes during retrieval. When viewing previously memorized scenes, those that were forgotten (misjudged as novel) still elicited stronger pupil constrictions than those correctly judged as familiar. Furthermore, pupil constriction was influenced more strongly if images were judged with high confidence. Thus, we propose that pupil constriction can serve as a marker of novelty. Since stimulus novelty modulates the efficacy of memory formation, our pupil measurements during learning indicate that the later forgotten images were perceived as less novel than the later remembered pictures. Taken together, our data provide evidence that pupil constriction is a physiological

correlate of a neural novelty signal during formation and retrieval of declarative memories for complex, natural scenes.

Introduction

Distinguishing novel from familiar items is a key function of the nervous system with immediate relevance for survival (Sokolov, 1963). Consequently, many parts of the human brain react differently to novel as compared to familiar items. For instance, several processes of memory encoding are only triggered if the item is considered novel (Knight, 1996), while familiar items trigger memory retrieval processes (Gonsalves, Kahn, Curran, Norman, & Wagner, 2005; Squire, Wixted, & Clark, 2007). Various brain areas are involved in distinguishing familiar from unfamiliar stimuli, including the medial temporal lobe (MTL; Gonsalves et al., 2005; Knight, 1996; Rutishauser, Mamelak, & Schuman, 2006; Rutishauser, Schuman, & Mamelak, 2008; Viskontas, Knowlton, Steinmetz, & Fried, 2006; Xiang & Brown, 1998), ventrolateral prefrontal cortex, and dorsal posterior parietal cortex (Desimone, 1996; Grill-Spector, Henson, & Martin,

Citation: 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):11, 1–20, <http://www.journalofvision.org/content/13/2/X11>, doi:10.1167/13.2.11.

2006; Kumaran & Maguire, 2009; Montaldi, Spencer, Roberts, & Mayes, 2006; Ranganath & Rainer, 2003; Yonelinas, Otten, Shaw, & Rugg, 2005). While some memories are implicit and only expressed in changed behavior, others are explicit such that observers can declare and assess the quality of their own memories (Tulving, 2002). The formation and retrieval of such declarative memories requires the MTL (Squire, Stark, & Clark, 2004). This differential engagement of particular structures such as the MTL has been used successfully to predict whether novel items will later be remembered (Chadwick, Hassabis, Weiskopf, & Maguire, 2010; Hassabis et al., 2009; Johnson, McDuff, Rugg, & Norman, 2009; McDuff, Frankel, & Norman, 2009; Polyn, Natu, Cohen, & Norman, 2005; Rutishauser et al., 2006) and whether, during retrieval, an item will be recognized correctly and with which confidence (Paller & Wagner, 2002; Rissman, Greely, & Wagner, 2010). These measurements require either invasive access to the brain or sophisticated machinery to infer brain activity (e.g., functional magnetic resonance imaging), severely limiting their applicability outside the laboratory and in clinical practice. Here we ask whether the pupil size, a simple and outwardly accessible physiological measure, can similarly serve as a probe of declarative memory processes.

Besides its rapid constriction in response to bright stimuli (Loewenfeld & Lowenstein, 1993), the pupil reacts to a variety of cognitive processes, such as attention (Daniels, Hock, & Huisman, 2009; Kahneman, 1973; Karatekin, 2004; Karatekin, Couperus, & Marcus, 2004), emotions (Bitsios, Szabadi, & Bradshaw, 2002; Charney, Scheier, & Redmond, 1995; Nagai, Wada, & Sunaga, 2002; Simpson & Molloy, 1971; Steinhauer, Boller, Zubin, & Pearlman, 1983), arousal (Bradshaw, 1967; Yoss, Moyer, & Hollenhorst, 1970), decisions (Simpson & Hale, 1969), or cognitive load (Hess & Polt, 1964). Cognitive processes can enhance or inhibit pupil constriction via several, mostly noradrenergic and cholinergic, pathways (Samuels & Szabadi, 2008). This results in the pupil size exhibiting complex temporal response patterns. In the context of memory, pupil size has long been known to increase with working memory load (Granholm, Asarnow, Sarkin, & Dykes, 1996; Kahneman & Beatty, 1966). Since the neural mechanisms of short- and long-term memory are largely distinct and interact in complex ways (Ranganath & Blumenfeld, 2005), it is unclear whether a similarly clear relationship between long-term memory performance and pupil size exists. Only very recently, several studies have started to address the relation between pupil size and forms of memory other than working memory (Heaver & Hutton, 2011; Kafkas & Montaldi, 2011; Otero, Weekes, & Hutton, 2011; Sterpenich et al., 2006; Vö et al., 2008). During memory retrieval of words presented visually or

auditory, the pupil dilates more for words that have been shown in a preceding study session (Otero et al., 2011; Vö et al., 2008). While only familiar words were used, such that no word itself was novel, these previous studies showed that the pupil can distinguish items based on whether they had recently been presented or not. Similarly, the pupil is found to be larger when showing a famous as compared to a nonfamous face, a difference possibly related to the recognition of the face (Maw & Pomplun, 2004). In all these studies at least part of the items (famous faces, all words) were previously known to the level of the particular exemplar. Hence, effects of novelty and memorization could not be studied in the same paradigm. A recent study showed that the pupillary response to natural stimuli is predictive of later memory during incidental memory encoding (Kafkas & Montaldi, 2011). It remains unclear, however, how these responses relate to a task where memory encoding is effortful rather than incidental. To the best of our knowledge, no study so far has addressed the relation of pupil size to the formation *and* retrieval of novel items (i.e., previously unseen exemplars of a potentially known category) in an explicit memory task, nor has any study used complex natural stimuli to probe the relation between pupil size and memory during retrieval.

In a series of four experiments, we monitored the pupil diameter during presentation of novel natural scenes, which we either asked observers to explicitly memorize (Experiments 1 through 3) or to merely use for performing a mock task (Experiment 4). We used these data to test whether pupil size was predictive of whether or not an image will be remembered during later retrieval. We continued by measuring the pupil diameter either during explicit retrieval, where observers were either asked to distinguish the previously presented scenes from novel ones (Experiments 1 through 3) together with a confidence rating, or during incidental retrieval (Experiment 4). We used this second dataset to test whether pupil size related to the subjective and/or objective novelty of a stimulus and to the confidence an observer has in their memories. Taken together, our experiments aim at linking pupil size to formation *and* retrieval of declarative memories, and at an outwardly accessible physiological marker of novelty.

Methods

Participants

A total of 48 volunteers participated in four experiments (age range: 18–26; 14 male, 34 female). Experiments 2 and 3 were performed by the same 16 participants, while Experiments 1 and 4 were per-

formed by disjoint sets of 16 individuals each. To ease notation, the term “individual” will hereafter refer to participant and session, such that there are 16 “individuals” per experiment and 64 “individuals” in total. For statistical analysis all these individuals will be treated as independent. All observers had normal or corrected-to-normal vision, were naive to the purpose of the studies, and gave informed written consent before each experiment. The experiments conformed to the ethical principles of the Declaration of Helsinki and were approved by the local ethics commission (Ethik-kommission FB04 Philipps-University Marburg).

Setup

Stimuli were presented on a 21" SyncMaster CRT screen (Samsung, Ridgefield Park, NJ) at a viewing distance of 70 cm. The refresh rate of the screen was 85 Hz and the resolution was 1280×1024 pixels. Stimuli were generated on an Optiplex 755 DELL computer, using Matlab (Mathworks, Natick, MA), the Psychophysics toolbox (Brainard, 1997), and EyeLink toolbox (Cornelissen, Peters, & Palmer, 2002) extensions. Pupil size was tracked with an infrared sensitive camera (EyeLink 2000, SR Research, Kanata, ON, Canada) with a sampling rate of 500 Hz. Observers' heads were supported by a chin- and forehead-rest. The eye-tracker was (re)calibrated before each experimental phase using a 13-point calibration grid.

To ensure that pupil diameter was not confounded by eye position due to foreshortening of the pupil relative to the EyeLink camera, we performed an independent experiment to verify the relation between measured eye position and pupil diameter. Despite large intrasubject variability we found the expected decrease of measured pupil size with eccentricity. However, for the most extreme eccentricity of relevance here, the positions of the stimulus corners, this effect was well below 2% (relative drop of measured pupil size due to eccentricity) for all tested luminance levels (0 cd/m^2 , 0.9 cd/m^2 , 4.7 cd/m^2 , 12.2 cd/m^2) and 1% on average over these luminance levels. Hence—in the light of the large variability of pupil size even for a uniform visual stimulus while fixating at constant eccentricity—the effect of eye position on the measured pupil diameter was negligible for the stimulus sizes considered in the present study. Hence, pupil diameter differences were not confounded by possible eye movements that observers were free to make during image presentation.

Stimuli

Three different image databases consisting of 200 images each were used (Figure 1A); images for

Experiments 1 and 3 used the same dataset as Rutishauser, Ross, Mamelak, and Schuman (2010); images for Experiment 2 were from the data set of Li, VanRullen, Koch, and Perona (2002); and Experiment 4 used different images from the same database, as well as some additional ones from various sources. The database for Experiments 1 and 3 contained images from the following categories: houses, general landscapes, vehicles, phones, and animals. The database for Experiment 2 had the categories: bushes, mountains, mushrooms, forests, and water; for Experiment 4: flowers, general landscapes, mushrooms, forests, and meadows. Images in Experiments 2 and 4 were chosen to be more alike per category than images in Experiments 1 and 3, and thus expected to be more difficult to memorize. In each experiment, all five categories were used equally often (20 images per category for memorization plus 20 for retrieval). All images were displayed as 384×256 pixels (width \times height) in size and rescaled to this size if necessary. This corresponded to about 10° by 6° of visual angle. In all experiments, images were presented on a gray background that had the same luminance as the blank preceding and following each image (4.7 cd/m^2).

Procedure

In Experiments 1 through 3, all observers were first asked to memorize 100 different images (*memorization phase*, Figure 1B shows example traces), each of which was presented for 1 s, followed by an 8-s blank period, during which a gray screen (4.7 cd/m^2) with a central fixation mark was presented. The memorization phase was followed by a distraction phase (5 min in Experiments 1 and 2; 20 min in Experiment 3). During the distraction phase observers performed a Stroop-task, which was merely to prevent active rehearsal and was not considered further. During the *retrieval phase* (Figure 1C shows example traces) the 100 “objectively familiar” images that had been presented during memorization were presented again, randomly intermixed with 100 different novel images. Each image was presented for 1 s and followed by a blank period, during which observers simultaneously indicated by pressing one of six buttons whether the image had been presented during the memorization phase (“familiar image”) or not (“novel image”) and how confident they were of their decision (three levels of confidence: “Confident,” “Probable,” or “Guess”). After a further 1.5-s blank period, the next image was shown; that is, once the response was given, the next trial started automatically, rather than the experiment being self-paced.

Experiment 4 was similar, with the exception that observers were not instructed to memorize images.

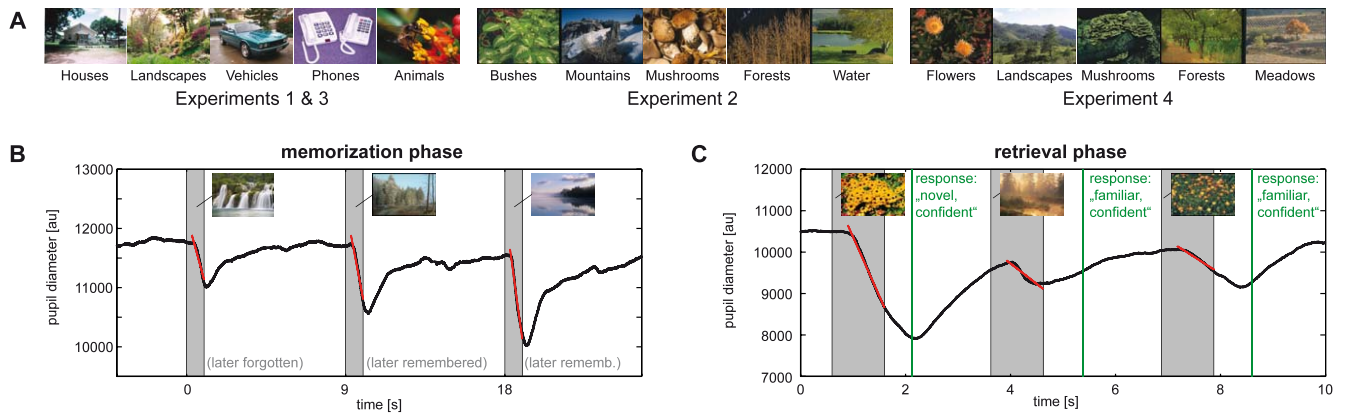


Figure 1. Stimuli, procedure, and example raw data. (A) Example stimuli for each image category and experiment. (B) Memorization phase. Each image was shown for 1 s (gray-shaded areas) followed by a 8-s blank period during which a central fixation mark was shown (white areas). The fixation mark was red for 3 s after image offset, indicating that observers should avoid blinks, and green otherwise. Black line shows corresponding pupil trace for one individual from Experiment 2. Constriction was stronger for the images later remembered compared to later forgotten images, as quantified the more negative slope during image presentation (red lines: linear fit to pupil trace from 300 ms after stimulus onset to stimulus offset). (C) Retrieval phase. Each image was presented for 1 s (gray-shaded areas), followed by a blank period during which a fixation mark was shown (white areas). During the blank, the observer had to respond by pressing one of six keys to simultaneously indicate their novel/familiar judgment and their confidence in their judgment (green lines). After a further 1.5 s-blank period, the next stimulus was onset; that is, once the response is given, control over the pace of the experiment is— not with the observer. The fixation mark was red for after-image offset till 0.5 s after the response, indicating that observers should avoid blinks, and green otherwise. Black line shows corresponding pupil trace for one individual (the same as panel (B)). Constriction was stronger for the novel image compared to the familiar images, as quantified by a more negative slope (red lines). Note the different time axis in panels (B) and (C).

Rather, they were exposed to them with the instruction to “familiarize” themselves with the images and to “become an expert” in order to qualify as “jury member for an art competition.” Since there was no incentive to memorize the images, no Stroop task was performed during the 5-min break. The retrieval phase was replaced by a rating phase, in which observers had to click on a continuous scale from 0 to 100 to rate how beautiful the image was they had just seen. Otherwise, the exposure phase was matched to the memorization phase of Experiments 1 through 3, and the rating phase to the retrieval phase.

To reduce image-specific effects as far as possible, in each of the experiments, half of the images were used for memorization in one half of observers, and the other half as novel images during retrieval, while this assignment was reversed for the other half of observers.

In all experiments, during memorization/exposure the blank screen contained a central fixation mark, which turned from red to green 3 s after image offset. During retrieval/rating, the mark was red from image offset to 0.5 s after the response. Throughout, observers were asked to fixate the mark while no image was shown, and to avoid blinks while the mark was red or the image was on. Observers controlled their eye blink behavior as instructed (see below).

Analysis

Performance

Performance during retrieval in Experiments 1 through 3 was defined using nomenclature from signal detection theory (SDT). An image correctly identified as familiar is considered a “hit” (“true positive,” “remembered”), an image incorrectly identified as familiar is a “false alarm” (“false positive”), a image incorrectly identified as novel is a “miss” (“forgotten”), and an image correctly identified as novel is a “correct reject.” To obtain a finer grained performance analysis per individual, subjective reports were sorted by confidence (in order of “familiar confident,” “familiar probably,” “familiar guess,” “novel guess,” “novel probably,” and “novel confident”), and the criterion was shifted along these confidence levels resulting in a 6-point receiver operating characteristic (ROC) curve per observer as a function of the confidence at which subjects rated trials as “familiar.” The confidence levels in the ROC are cumulative and are calculated according to standard SDT analysis methods as follows (Macmillan & Creelman, 2005). The first data-point of the ROC (lower left corner) shows the memory performance only for trials that were rated “familiar confident,” and the second for all trials of the previous data-point plus those rated “familiar probably.” This

process continues till the last data-point where all confidence ratings are considered as “familiar,” resulting in a true positive rate of 1 (by definition). The degree of asymmetry of the ROC of each individual subject was assessed by fitting the z-transformed ROC with a straight line (using least-squares regression). The slope of this line is 1 if the ROC is symmetric and significantly less than 1 if it is asymmetric (Macmillan & Creelman, 2005; Ratcliff, Sheu, & Gronlund, 1992). A slope of less than 1 confirms that MTL-dependent declarative memories were formed (Wais, Wixted, Hopkins, & Squire, 2006).

Pupil size

Pupil diameter was measured by the Eyelink software and used at its original resolution of 500 Hz without additional filtering throughout, both in figures and for analysis. Periods of blinks, as detected by the Eyelink software, were interpolated using cubic spline interpolation (see Figure 1B and C for example traces). Of the total measurement time during memorization/exposure, blinks accounted for 8.3% of data points. This number dropped to 0.8% during image presentation and to 2.1% for the 3-s after-image offset when observers were instructed not to blink by the red fixation mark. This instruction was effective, as observers blinked during 21.1% of the remaining time. During retrieval/rating, blinks accounted only for 3.7% of the overall measurement time, which reduced to 0.3% during image presentation, was 3.7% while blinking should be avoided during the blank (till 0.5 s after response) and nearly doubled (7.2%) while observers were allowed to blink. As such, the overall low numbers of blinks, in particular during stimulus presentation and directly thereafter, showed that observers complied with instruction and that blinks were not a critical factor in the analysis. To facilitate inter-subject comparison, the entire pupil data of each individual was z-scored in each experimental phase (i.e., separately for memorization and retrieval). Note that with the exception of mean traces, analysis was not affected by this normalization and no per image or per trial normalization was performed.

Pupil slope

We used the slope of the pupil trace following stimulus onset to quantify the pupil response for each trial. The slope of the pupil trace was determined by fitting a line (using linear regression, best fit in a least-squares sense) to the pupil size trace as a function of time from 300 ms after image onset till image offset (Figure 1B and C). Out of the conceivable measures that summarize the pupil trace in a single number, we chose this slope for the following reasons: first, the

slope quantifies directly the initial pupil constriction in each trial, while later differences are likely a mixture of different effects; second, data before image offset allows a direct comparison between memorization and retrieval phase, as no motor response occurred during the presentation itself in either phase, third, the slope is insensitive to slow drifts of pupil size over the course of the experiment, and fourth, the slope is not affected by trial-specific shifts of the pupil size, related to overall vigilance, attention, or alertness. For the negative slopes typically observed, a steeper (more negative) slope implies stronger pupil constriction.

Statistics

Pooled pupil traces were compared point-wise using *t* tests (two tailed, unpaired). Comparing data across individuals used paired two-tailed *t* tests when the effect sizes are of interest; to test whether a significant fraction of observers showed the same effect direction irrespective of size, sign tests (two sided) were used. For all comparisons, both sign-test and *t*-test results are reported, whereas significance markers in figures refer to sign test only.

Signal Detection Theory

Trial-by-trial prediction of pupil slopes was analyzed by computing the ROC curve for discriminating misses from hits based on pupil slopes. The area under the ROC curve (AUC) was used for quantification, with 50% being chance and 100% being perfect discrimination.

Terminology

The terms “objectively familiar” and “objectively novel” denote images shown during retrieval that have and have not been shown during the memorization phase, respectively. The terms “subjectively familiar” and “subjectively novel” refer to judgment of the observer during retrieval, irrespective of objective familiarity. Where appropriate, SDT terminology is used in addition as described above (*hit*: subjectively and objectively familiar; *false alarm*: subjectively familiar, but objectively novel; *correct reject*: subjectively and objectively novel; *miss*: subjectively novel, but objectively familiar).

We will consistently use the terms “stronger constriction” and “weaker constriction” throughout the Results section because the pupil size changes were, partially because of pupil-light responses to stimulus onset, typically in the direction of reduced pupil size (smaller pupils). As we measured pupil size by video-oculography, we could not distinguish constriction from dilation in a physiological sense. That is,

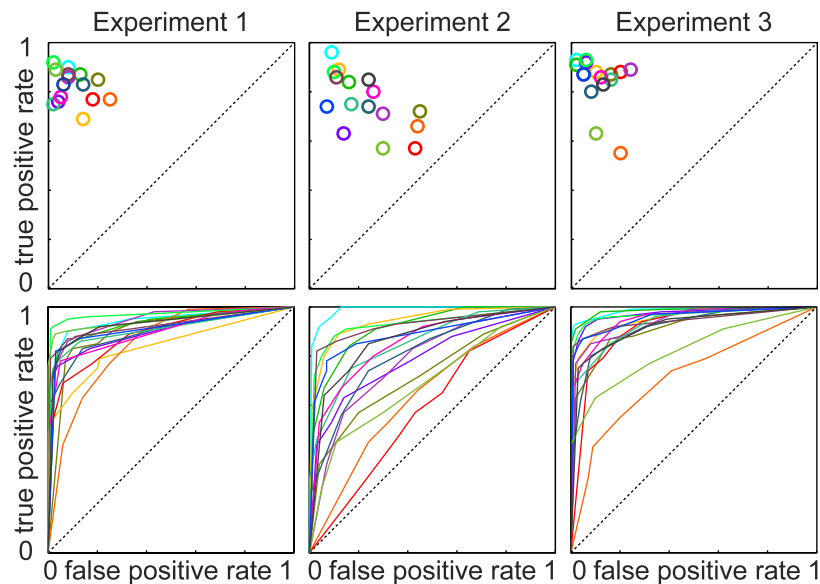


Figure 2. Memory performance. Top row: Each data point represents the performance of one individual in terms of the fraction of true positives (images correctly judged as familiar divided by total number of familiar images) as a function of the fraction of false positives (images incorrectly judged as familiar divided by the total number of novel images). Bottom row: 6-point ROC curve for each subject, as a function of confidence. Note that the data points of the top row lie on the corresponding curve. Each column represents one experiment; color-code identifies individuals. Individuals share the same color code in this figure.

“stronger/weaker constriction” could also mean a “weaker/stronger dilation” on top of an overall pupil size reduction. For clarity of presentation we, however, opted to use the term constriction throughout.

Results

Sixty-four individuals (48 distinct participants, see Methods) performed a total of four experiments to test the relation between pupil size and memory encoding and retrieval. Experiments 1 through 3 tested explicit memorization with different degrees of difficulty, and Experiment 4 tested incidental memorization. Performance and its relation to pupil size was thus tested for Experiments 1 through 3, while for Experiment 4, only the effect of objective novelty, regardless of the unknown subjective judgment of novelty, was tested.

Performance (Experiments 1 through 3)

In Experiments 1 through 3, 16 observers performed a memorization task. As expected by design, recall performance differed between Experiments 1 and 3 on the one hand, and Experiment 2 on the other hand [percentage correct: Experiment 1: $86.3\% \pm 5.2\%$; Experiment 2: $76.7\% \pm 11.1\%$; Experiment 3: $86.6\% \pm 7.2\%$, $F(1, 2) = 7.51$, $p = 0.0015$], with pronounced differences both in the false positive rate [images that

were mistakenly classified as previously seen, $9.8\% \pm 6.9\%$; $22.6\% \pm 12.9\%$; $10.1\% \pm 6.7\%$, $F(1, 2) = 9.41$, $p < 0.001$] and small differences in the true positive rate [images that were correctly classified as previously seen, $82.3\% \pm 6.4\%$; $76.1\% \pm 11.6\%$; Experiment 3: $84.2\% \pm 10.5\%$, $F(1, 2) = 3.0$, $p = 0.058$, Figure 2, top]. However, in all experiments, all observers’ performance was clearly above chance (50%) and significantly so, $t(15) > 9.6$, $p < 10^{-6}$ for all experiments.

All individuals in all experiments had “confident” and “guess” trials for “familiar” and “novel” responses, but for three individuals there were no “probable” trials for at least one judgment (familiar or novel). In total, “confident” was chosen in $61.9\% \pm 19.1\%$ of trials, “probable” in $18.9\% \pm 10.0\%$ and “guessing” in $19.2\% \pm 12.7\%$ of trials.

Using these subjective confidence judgments, we computed a six-point behavioral ROC curve for each individual (Figure 2, bottom). Observers had a good sense of the quality of their memories as performance clearly dependent on the levels of confidence indicated. The z-transformed versions of these ROC curves had slopes clearly below 1 in 44 out of 48 individuals (mean $\pm SD$ of slope: 0.68 ± 0.24 ; t test for difference to 1: $t[47] = 9.36$, $p < 0.001$). Such slopes below 1 indicate ROC curve asymmetry, which in turn is typically observed for declarative memories (for review see Squire et al., 2007; Wixted, 2007). These data confirmed the validity of our task as a test of declarative memory.

Pupil constriction during memorization predicts later retrieval success (Experiments 1 through 3)

About 300 ms after the onset of an image, the pupil started to constrict (Figures 1B and 3A through C). This constriction was partially due to the image onset as such and the somewhat higher luminance of the images (Experiments 1 and 3: 8.2 ± 4.4 cd/m²; Experiment 2: 4.8 ± 3.3 cd/m²; Experiment 4: 5.5 ± 3.3 cd/m²) as compared to the blank background (4.7 cd/m²). However, even in Experiment 2, for which blank luminance and average luminance were closest, there was still a pronounced constriction (Figure 3B). More importantly, however, when comparing the pupil response to images that an observer later remembered from those that the observer did not remember later, clear differences were visible from about 600 ms after stimulus onset. Images that were later forgotten (Figure 3A through C, red) elicited less constriction than images that were later remembered (Figure 3A through C, blue). Pooled over all observers of each experiment, the difference between the traces became significant (at an uncorrected 5% alpha level) 629 ms (Experiment 1), 802 ms (Experiment 2), and 962 ms (Experiment 3) after stimulus onset and remained so for an extended period of time (Figure 3A through C, gray bar; some isolated “significant” time points occur earlier in Experiment 1). When correcting for multiple comparisons (as there are 1,500 samples considered per trial), significant differences occurred 656 ms (Experiment 1) and 823 ms (Experiment 2) at an expected false discovery rate (FDR; Benjamini & Hochberg, 1995) of 5%, with a similar overall pattern that did not reach corrected significance for Experiment 3. This showed that already the mean traces had the tendency to distinguish later remembered from later forgotten items.

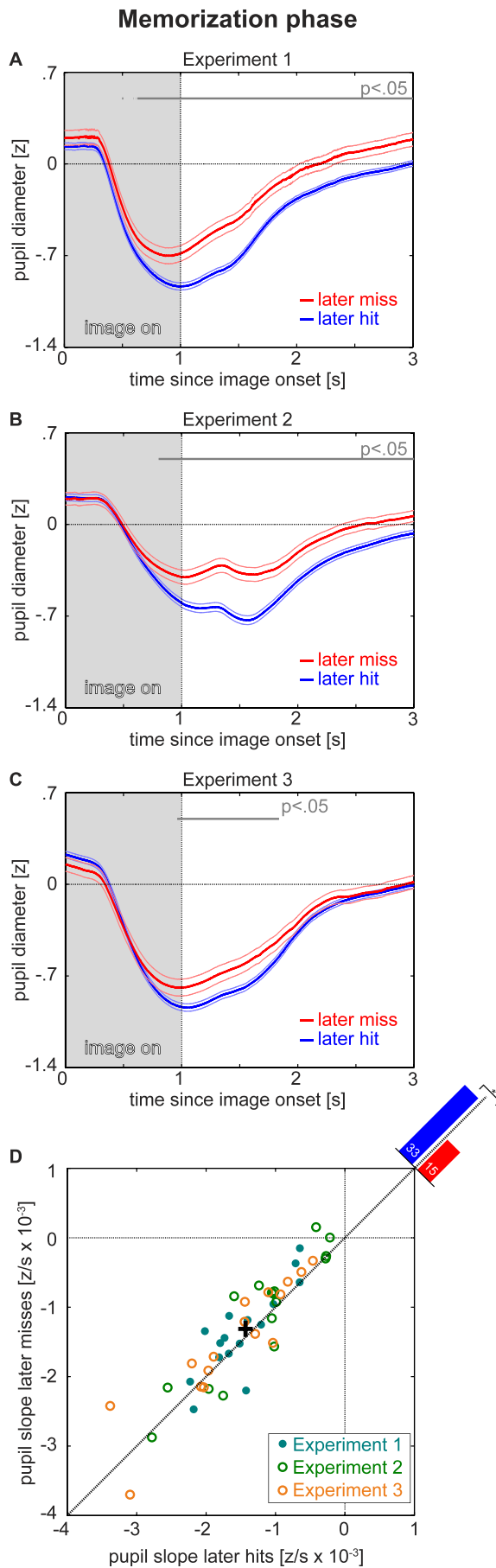
Although the pooled data considered so far provided an initial hint at the difference between later remembered and later forgotten images, differences within each individual were more telling, especially if the interobserver differences in pupil response was comparably large. To capture a single trial’s pupil trace in a single number, we used the slope for each image presentation, starting 300 ms after image (trial) onset and ending at image offset (Figure 1B, C). Across all experiments, more than two thirds of individuals (33/48, Figure 3D) showed steeper (more negative) slopes for images later remembered than for those later forgotten (Figure 3D). This fraction was significant, even when effect sizes in the individuals were ignored ($p = 0.01$, sign test); and the tendency was consistent across all experiments (Experiment 1: 12/16; Experiment 2: 10/16; Experiment 3: 11/16). If effect sizes were considered and mean slopes for remembered and forgotten were compared, significance prevailed, $t(47) =$

2.36 , $p = 0.02$, paired t test. Hence, in a significant majority of individuals and on average, the slope during image presentation for an image to be memorized was steeper for later remembered than for later forgotten images. Thus, the slope during memorization predicted whether an image would later be remembered or forgotten.

Instead of considering averages of the single trials’ slopes, SDT analysis allows testing the trial-by-trial prediction of whether an individual image will be remembered or forgotten. Hence for each individual we computed the ROC curve for distinguishing these alternatives based on the slope during image presentation in the memorization phase and quantified them by the AUC. The AUC was larger than chance (50%) in 32/48, and on average reached $53.3\% \pm 8.8\%$, which was slightly, but significantly, $t(47) = 2.57$, $p = 0.01$, above chance. This verified that the slope alone was a weak, but significant, predictor of later retrieval success. It seems likely that using other features of the pupil trace rather than simply the slope, or adapting parameters for each individual (e.g., time window in which the slope is computed), could improve prediction performance. In light of this, it is important to note that this very simple measure with no adaptation to individual differences already achieved significant above-chance single-trial prediction performance. This is an indication of the robustness of the pupil signal across individuals, datasets, and conditions.

Prestimulus baseline (pupil size at image onset) does not predict memorization (Experiments 1 through 3)

Since generic (i.e., image-independent) cognitive factors, such as attention or arousal, can potentially affect memorization and pupil size, it is conceivable that pupil size prior to stimulus onset is already predictive of later memorization. In addition, if pupil size should recover differentially to baseline for later hit as compared to later miss trials, pupil size at trial onset could still be influenced by the preceding trial. This could, in combination with a possible serial dependence for memorization performance, account for part of the observed results. We therefore performed the same analyses as for the slopes for the prestimulus baseline, which we define as the pupil size at image onset. In line with earlier observations (Kafkas & Montaldi, 2011), we found no significant effects in any of the experiments, no matter whether effect sizes were considered (Experiment 1: $t[15] = 0.46$, $p = 0.65$; Experiment 2: $t[15] = -1.23$, $p = 0.24$; Experiment 3: $t[15] = 0.72$, $p = 0.49$; all experiments: $t[47] = -0.21$, $p = 0.83$) or not (Experiment 1: 8/16 larger prestimulus baseline for misses; Experiment 2: 9/16; Experiment 3: 7/16; all



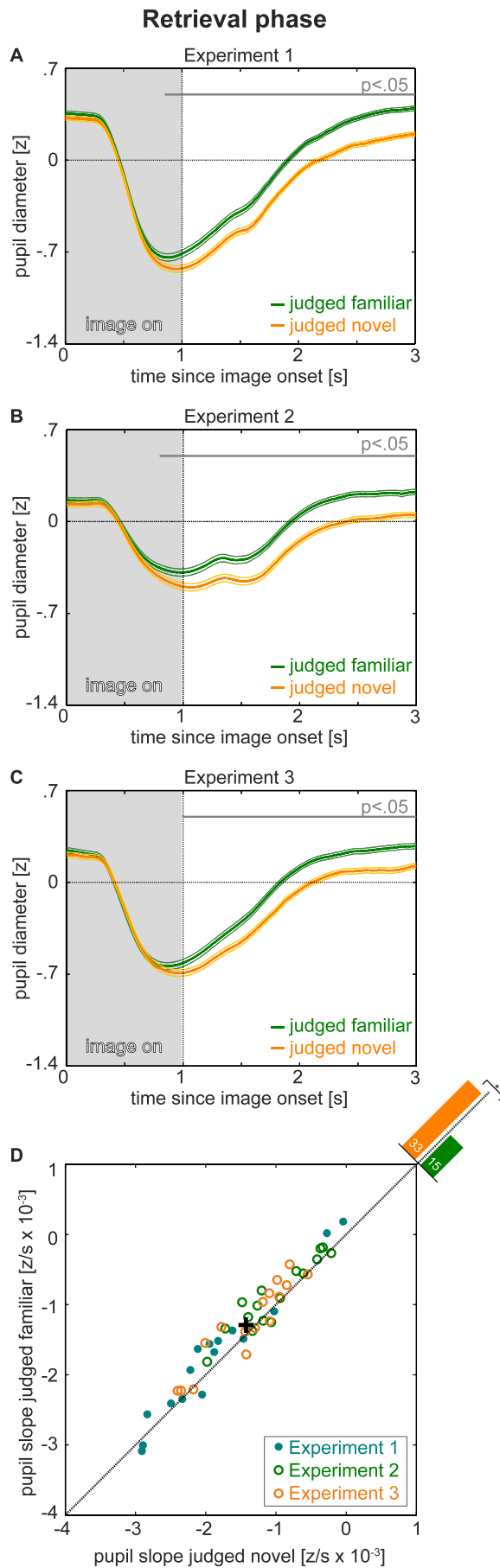
experiments: 24/48, $p = 1$, sign test). While this result alone cannot fully exclude that generic factors contribute to memorization and pupil dilation alike, it rules out that the observed effects on slopes are an artifact of incomplete recovery of the previous trial.

Image luminance does not predict memorization (Experiments 1 through 3)

Since luminance is an obvious factor in driving pupil constriction, we next verified that image luminance was not a common factor in causing larger constrictions as well as better memory. We found that there was no significant difference in the luminance of remembered compared to forgotten images. When considering individuals, only 3 out of 48 showed “significance” at the 5% level. Due to the large number of individual tests, such “significance” was of little meaning, and more importantly, two of those three showed higher mean luminances for forgotten images, and one higher luminance for images later remembered. Similarly, no significant effect was seen when comparing the mean luminances for forgotten and remembered images across individuals, $t(47) = 0.72$; $p = 0.47$. We thus concluded that the increased constriction seen for later remembered images was not due to higher luminance of these images. Although at this point other possible image-related factors driving pupil and memory alike are still possible, a reversal of this effect during retrieval (see below) will show that there is no low-level image feature possibly responsible for the observed effects.

←

Figure 3. Response of the pupil during memorization. (A through C) Average pupil trace pooled over all observers of each experiment for images that were later remembered (hits, blue) or forgotten (misses, red). Mean and standard error over trials (i.e., not weighted by observer); gray bars indicate significant time points at $p < 0.05$ (A) Experiment 1, (B) Experiment 2, (C) Experiment 3. (D) The average slope of constriction (see Figure 1B for single trial examples) during the 1-s long image presentation was steeper for later remembered images (hits, x-axis) compared to later forgotten images (misses, y-axis). Note that slopes are in general negative, thus steeper slopes imply more negative values; that is, stronger constriction corresponds to more negative values. Each data point corresponds to one individual in one experiment. Points above the diagonal imply later hits have a more negative (steeper) slope on average for the respective observer than later misses. Black cross denotes average over observers and standard error in each dimension. Bars in upper right corner denote the number of points above and below the diagonal. Significance marker at bars denotes sign-test result, while in the text paired t tests (considering effect sizes) and sign tests (not considering effect sizes) are reported.

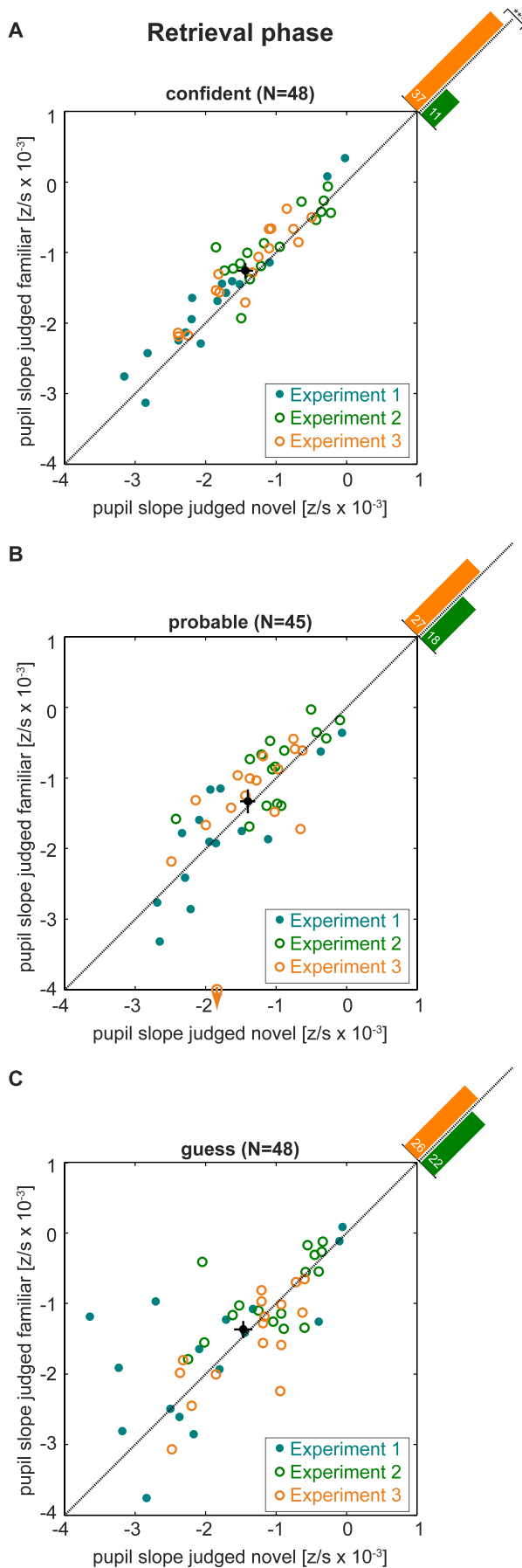


Pupil constriction during retrieval distinguishes familiar from novel stimuli (Experiments 1 through 3)

Analysis so far considered the memorization phase. Was the pupil also informative of whether or not an image was considered familiar or novel during retrieval? During retrieval, observers saw 200 images, half of which they had seen during memorization, half of which were novel (randomly intermixed). First, we compared the subjective judgment, that is, images for which observers responded “familiar” to those to which they responded “novel” (regardless of whether this judgment was correct or incorrect). When pooling data within experiments, subjectively familiar images showed a weaker constriction compared to subjectively novel images (Figure 4A through C). The difference between the curves was significant (at an uncorrected 5% level) about 800 ms after image onset (uncorrected 5% level: Experiment 1: 852 ms; Experiment 2: 800 ms; Experiment 3: 992 ms; 5% FDR: 881 ms, 843 ms, and 1013 ms, respectively) and remained so till after the response. This was a first indication that novel stimuli evoked a stronger pupil constriction than familiar stimuli.

Similar to the previous analysis of the memorization phase, we quantified the individual trials by the slope of the pupil during image presentation. We found that for 33 out of 48 individuals, the slope was steeper (more negative) for unfamiliar as compared to familiar images (Figure 4D; $p = 0.01$ sign test; Experiment 1: 10/16; Experiment 2: 12/16; Experiment 3: 11/16) and a significant difference for the average, $t(47) = 4.71$, $p < 0.001$, black cross in Figure 4D. Thus, pupil constriction was stronger for images that were judged as novel compared to images that were judged as familiar.

Figure 4. Response of the pupil during retrieval according to the subjective judgment of familiarity. (A through C) Average pupil trace pooled over all observers of each experiment for images judged as novel (orange) or familiar (green). Mean and standard error over trials (i.e., not weighted by observer); gray bars indicate significant time points at $p < 0.05$. (A) Experiment 1, (B) Experiment 2, (C) Experiment 3. (D) The average slope of constriction during the 1-s image presentation was larger for stimuli judged novel (x-axis) compared to stimuli judged familiar. Each data point corresponds to one individual in one experiment. Points above the diagonal imply novel images have a more negative (steeper) slope on average for the respective observer than familiar ones. Black cross denotes average over observers and standard error in each dimension. Bars in upper right corner denote the number of points above and below the diagonal. Significance marker at bars denotes sign-test result.



Pupil constriction during retrieval is not related to the response time in the preceding trial

During retrieval, observers responded to the stimuli about 2 s after stimulus onset (Experiment 1: 1821 ms \pm 382 ms; Experiment 2: 1958 ms \pm 358 ms; Experiment 3: 1982 ms \pm 351 ms; median response times in each individual, mean and *SD* over individuals), with a slight (103 ± 170 ms) but significant advance for “familiar” responses, $t(47) = 4.17$, $p < 0.001$. Since the time between subsequent trials depended on response time (Figure 1C), we tested whether the response time in one trial had an effect on the pupil slope in the *subsequent* trial. At an uncorrected $p < 0.05$ level, we found significant correlations between slopes and preceding reaction times in only 6/48 individuals (Experiment 1: 1/16; Experiment 2: 1/16; Experiment 3: 4/16). This is still in the range expected for spurious correlations, and indeed no significant correlations were found at a Bonferroni-corrected level of 0.05/48, and neither at a less conservative threshold given by an expected FDR of 5% according to the Benjamini and Hochberg (1995) method. This renders it unlikely that the slope results were confounded by incomplete return to baseline or by any other effect related to distinct reaction times in the preceding trial.

Effect of familiarity on pupil constriction depends on high confidence (Experiments 1 through 3)

Above, we observed that pupil constriction was stronger for subjectively novel images. Does this effect depend on the confidence of the judgment? To address this question, we repeated the analysis split by confidence level. That is, we compared trials for which observers reported “confident familiar” to trials “confident novel” (Figure 5A), “probably familiar” to “probably novel” (Figure 5B) and “guess familiar” to

←
Figure 5. Dependence of the pupil response during retrieval on confidence. The slope of constriction differed more prominently between stimuli judged familiar and novel when observers indicated a high confidence (see text for statistics). Plot analogous to Figure 4D, but split by confidence levels confident (panel A), probable (panel B) and guessing (panel C). Note that in (B) there are only 45 datapoints, as three observers did not have data for all confidences. Outlier depicted with arrow in panel B has coordinates $(-0.0018, -0.0068)$, which falls well outside the depicted range in the vertical dimension. In all panels, p values and significance markers at bars denote significant test results.

“guess novel” (Figure 5C). For the highest degree of confidence, a significant effect was observed, which was somewhat more prominent than for all data considered (stronger constriction for novel in 37/48 cases, sign test: $p < 0.001$; t test on averages: $t[47] = 4.84$, $p < 0.001$). However, for the confidence level “probable,” for which only 45 cases were available, as well as for the level “guessing,” significance vanished (probable: 27/45, $t[44] = 0.16$, $p = 0.87$; guessing: 26/48, $t[47] = 0.95$, $p = 0.35$). If only correct trials were considered, the pattern did not change qualitatively (confident: 35/48, $t[47] = 4.75$, $p < 0.001$; probable: 28/45, $t[44] = 0.14$, $p = 0.89$; guessing: 26/47, $t[46] = 0.21$, $p = 0.84$). Hence the absence of a significant effect for lower confidence is not due to more errors in these trials. To directly compare the “confident” level with the other two levels combined, we defined the effect size for each individual as the difference between familiar and unfamiliar mean slopes. Effect sizes were larger for the “confident” level in 33/48 individuals ($p = 0.01$, sign test). Taken together, these data indicated that certainty about one’s familiarity judgment is critical for the relation between novelty and pupil constriction.

Correct and incorrect trials (Experiments 1 through 3)

So far we compared trials according to the subjective judgment of familiarity made by observers, regardless of whether this judgment was correct or incorrect. For correctness of judgment itself we found no influence on pupil constriction: neither on average, $t(47) = 1.17$, $p = 0.09$, nor when effect sizes were ignored (30/48 individuals with stronger constriction for incorrect judgments, $p = 0.11$, sign test). When only correct trials were considered, the results on familiarity were very similar to all trials: In 33/48 individuals constriction was stronger for novel compared to familiar images, $p = 0.01$, sign test; $t(47) = 4.67$, $p < 0.001$, t test). In contrast, when only considering incorrect trials, no significant effect of subjective familiarity was observed (26/48 individuals with stronger constriction for images judged novel, $p = 0.67$, sign test; $t[47] = 0.76$, $p = 0.45$, t test). While this might in part be a consequence of confidence being lower for incorrect judgments, it also motivates two additional pieces of analysis. First, does objective familiarity (ground truth) alone relate to pupil size, regardless of subjective judgment; second, if only objectively familiar images are considered, does the pupil still signal a difference between those judged as familiar and those misjudged as novel (i.e., comparing remembered and forgotten images)?

Objective familiarity (Experiments 1 through 4)

To address the first question and to allow inclusion of the incidental memory in Experiment 4, we next analyzed trials according to their objective familiarity (ground truth) alone, regardless of subjective judgment. Thus, we compared the response between images that were objectively familiar with images that were objectively novel, regardless of whether the observers indicated so. Considering Experiments 1 through 3, a similar picture emerged as for the subjective judgment: 33/48 ($p = 0.01$, sign test) and the average was significantly different, $t(47) = 3.86$, $p < 0.001$ (Figure 6A, blue, green, and orange).

In Experiment 4, we had subjects view images with neither an explicit instruction to memorize them nor with an explicit instruction to retrieve them. Instead, the instruction was to make a judgment on aesthetics (see Methods). Since observers made no explicit memory judgments, no analysis was possible to verify whether first exposure predicted later memory. We could, however, compare objectively novel against familiar images. Eleven out of 16 individuals had steeper slopes (stronger constriction) for novel images compared to familiar images, and the average difference between familiar and novel was on the edge of significance for Experiment 4 alone, $t(15) = 2.08$, $p = 0.055$. There was no confound by aesthetics as putative common factor, as familiar and novel images received similar ratings, $t(15) = 0.47$, $p = 0.64$. Thus, a difference between previously seen and novel images was apparent even in the absence of an explicit instruction to remember. Across all four experiments (Figure 6A), the effects of objective familiarity were highly significant, both for the fraction of observers that had steeper slopes for novel images (44/64) as well as for the average, $t(63) = 4.30$, $p < 0.001$. Taken together, this indicated that explicit memorization was not required to result in a stronger constriction for novel compared to familiar images.

For objectively familiar images, subjective novelty induces stronger constriction (Experiments 1 through 3)

For Experiments 1 through 3, we could directly compare the subjective judgment of familiarity to the objective familiarity of a stimulus. We therefore considered only those 100 images that had been shown during memorization (i.e., are objectively familiar) and compared the pupil slopes for those that the observer correctly considered familiar to those the observers incorrectly considered as novel (i.e., they forgot them). In 30/48 individuals, constriction was stronger for misses (images that were mistakenly judged as novel)

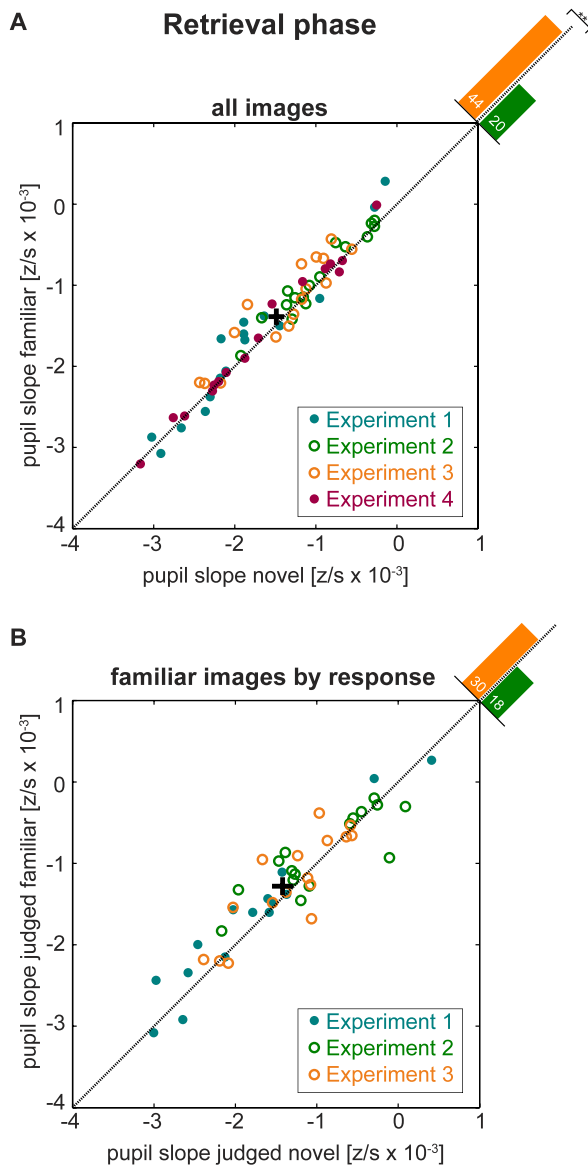


Figure 6. Pupil constriction during retrieval distinguishes between forgotten and remembered images (objective familiarity). (A) Pupil constriction was larger for objectively novel compared to objectively familiar images, regardless of subjective report (Figure analogous to Figure 4D), thus depicts ground truth regardless of subjective report. (B) Comparison of pupil constriction for remembered versus forgotten images during retrieval. Only objectively familiar images were considered. Pupil constriction was larger for forgotten compared to remembered images. x-axis: images misjudged as subjectively novel (forgotten), y-axis: images correctly judged as familiar. Note that the x-axis therefore corresponds to misses (familiar image judged novel) and the y-axis to hits (familiar image judged familiar), that is, axes are reversed relative to Figure 3D. More points (30/48) are above the diagonal and on average constriction is indeed significantly larger for misses (images judged as novel) than for hits (images judged as familiar), which indicates a reversal between hits and misses for retrieval as compared to memorization (compare to Figure 3D). This reversal also rules out that our findings are a result of image-specific features (e.g., luminance).

than for hits (images correctly judged as familiar). Although this fraction was not significant when effect sizes were ignored ($p = 0.11$, sign test), on average constriction was significantly larger for image misjudged as novel than for those judged as familiar, $t(47) = 2.42$, $p = 0.02$ (black cross in Figure 6B). This implies that for images that were in fact all familiar, those judged as novel evoked a stronger constriction than those correctly judged as familiar.

This result is particularly remarkable when compared to the data of the memorization phase (Figure 3D), which in each individual considered the exact same images with the same split in forgotten and remembered. During memorization, images that were later remembered had steeper slopes than those later forgotten. During retrieval this pattern reversed (Figure 6B): Those images that were forgotten (and thus mistakenly judged as unfamiliar) had steeper pupil slopes than those remembered (judged correctly as familiar). This gives rise to the idea that pupil constriction signals novelty in both cases; under this hypothesis images judged as less novel during memorization have shallower slopes and are less well remembered later, consistent with the data during memorization. Importantly, this finding also served as an additional control for image luminance and in fact any possibly confounding image feature—the fact that the identical images showed reversed patterns of pupil constriction during the different phases (memorization/retrieval) ruled out that memorization and pupil were driven predominantly by a common, image-specific factor.

Prestimulus baseline during retrieval

As for the memorization phase, we checked the effects of the prestimulus baseline (pupil size at trial onset) on the familiarity judgment. Performing the same analyses for pupil dilation at stimulus onset as for the slopes, we found no significant effects for individual experiments, no matter whether effect sizes were considered (Experiment 1: $t[15] = 1.27$, $p = 0.22$; Experiment 2: $t[15] = 1.62$, $p = 0.13$; Experiment 3: $t[15] = 1.18$, $p = 0.26$) or not (Experiment 1: 9/16 observers with larger prestimulus baseline for images judged as familiar; Experiment 2: 10/16; and Experiment 3: 8/16). When aggregating data over all experiments and ignoring effect sizes, there was still no effect of prestimulus pupil baseline on familiarity judgments (27/48, $p = 0.47$, sign test), but if effect sizes were taken into account a slight, but significant, difference was visible: The prestimulus baseline was larger for stimuli judged as familiar than for those judged as novel, $t(47) = 2.36$, $p = 0.02$. Whether the stimulus in the trial that follows is objectively familiar or novel cannot be

known beforehand. Hence, this difference in the prestimulus baseline needs to relate to the correctness of the following response. Indeed, for correct judgments the prestimulus baseline was larger for familiar than for unfamiliar stimuli in 36/48 individuals ($p = 7 \times 10^{-4}$, sign test; Experiment 1: 13/16; Experiment 2: 10/16; Experiment 3: 13/16) and this difference was significant also when effect size was considered, $t(47) = 3.66$, $p < 0.001$. When analyzing only incorrect trials, there was no difference between familiar and unfamiliar judgments, $t(47) = 0.17$, $p = 0.87$ with 28/48 having a larger prestimulus baseline, $p = 0.31$, sign test. When analyzing only correct trials, there was some effect on average, $t(47) = 2.33$, $p = 0.02$, but only a slight fraction of individuals that had a higher prestimulus baseline for familiar images, 29/48, $p = 0.19$, sign test. Thus, a higher prestimulus baseline goes along with correct responses with a possible slight benefit for responding correctly to familiar images. Since errors on novel images are rare as compared to errors on familiar images (there are fewer false alarms than misses, i.e., more correct rejects than hits), the residual familiar/novel difference for correct judgments may, however, relate to a ceiling effect. In any case, the prestimulus baseline during retrieval is more tightly coupled to correct/incorrect judgments than pupil constriction during image presentation (i.e., pupil slope) and in turn, this constriction is more clearly signaling novelty.

Relation between average slopes and performance (Experiments 1 through 3)

So far we have considered for each observer the relation of pupil slopes to images later remembered or forgotten (memorization) or familiar/novel (retrieval). While the effects were robust across individuals, there was substantial interobserver variability that remained unexplained. To test for interindividual effects, we related the average pupil slope of each observer to their memory performance (percentage correct). For average pupil slopes during memorization we found a significant correlation with performance when considering all images, $r(47) = -0.55$, $p = 5 \times 10^{-4}$, implying that steeper (more negative) slopes are related to better performance. Since this result could still be confounded by the fact that better performance means more “later remembered” trials and thus more trials with steeper slopes in the individuals with good memory, we analyzed data separately for hits and misses in each individual. For both we found significant correlations (hits: $r[47] = -0.54$, $p = 0.0001$, Figure 7A; misses: $r[47] = -0.51$, $p = 0.0002$, Figure 7B). These results also hold for Experiment 2 (hits: $r[15] = -0.62$, $p = 0.01$; misses: $r[15] = -0.54$, $p = 0.03$) and Experiment 3 (hits: $r[15] = -0.60$, $p = 0.01$; misses: $r[15] = -0.64$, $p = 0.007$) in

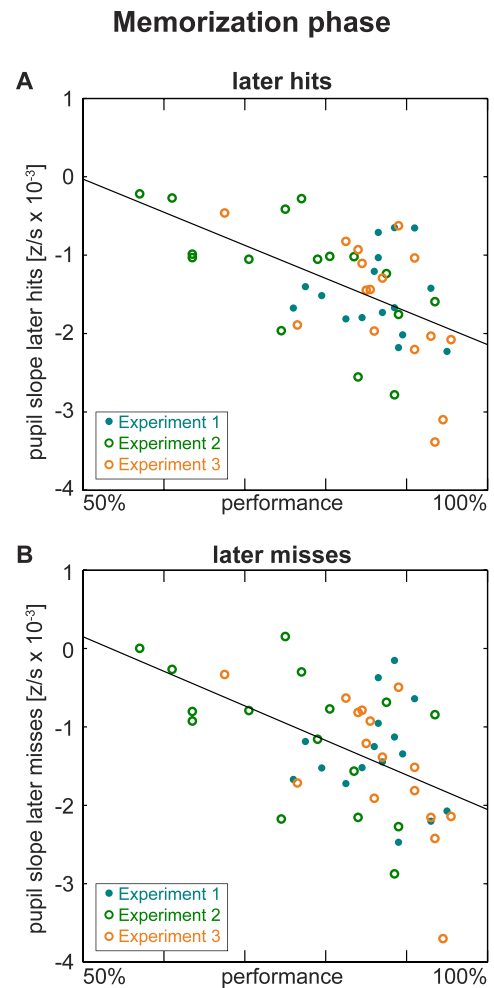


Figure 7. Relation between performance and average pupil slope. Average pupil slope during memorization plotted against individual performance. Each data point corresponds to one individual, colors indicate experiment number, regression line in black. (A) Later remembered images (later hits), (B) later forgotten images (later misses).

isolation, though not for Experiment 1 (hits: $r[15] = -0.03$, $p = 0.92$; misses: $r[15] = -0.07$, $p = 0.79$). Slightly weaker, but still significant, correlations were found for pupil slope during retrieval (all: $r[47] = -0.39$, $p = 0.006$; judged familiar: $r[47] = -0.37$, $p = 0.001$; judged novel: $r[47] = -0.41$, $p = 0.004$; objectively familiar: $r[47] = -0.37$, $p = 0.01$; objectively novel: $r[47] = -0.41$, $p = 0.004$). For Experiment 2, these correlations were significant (at $p < 0.05$) individually with the exception of objectively familiar, $r(15) = -0.49$, $p = 0.052$, while for the other experiments individually these correlations failed to reach significance. Since performance was higher in Experiments 1 and 3 than in Experiment 2, this difference is likely a ceiling effect on performance. In sum, these data show that across all experiments stronger pupil constriction was in general associated with better memorization performance.

Thus, individuals with better performance tended to have steeper slopes on average.

Discussion

We demonstrate that pupil size contains information about declarative long-term memory processes—both during memorization and retrieval—for a variety of natural stimulus sets and difficulty levels. We intentionally used realistic natural images with variability in luminance and other features, but our results could not be explained by image-specific features. During memorization, pupil constriction was stronger for images that were later remembered correctly. During retrieval, novel items triggered stronger constrictions than familiar items. Also, we observed a difference between familiar and novel images even in the absence of explicit instructions to memorize or retrieve. During retrieval, pupil constriction was stronger for images that were forgotten (i.e., misjudged as novel) compared to images correctly remembered (i.e., judged as familiar). Based on these results, we conclude that the extent of pupil constriction signaled how subjectively novel a stimulus is. While pupil constriction is clearly driven by increases in image luminance, the novelty component nevertheless modulated the extent of constriction additionally irrespective of luminance or any other image feature.

In learning, the subjective judgment of novelty is hypothesized to be a strong driver of memory formation (Kishiyama, Yonelinas, & Lazzara, 2004; Knight, 1996; Lisman & Grace, 2005). For example, items that are distinct and thus novel are later remembered better even if the novel attribute is task irrelevant (the von Restorff effect, von Restorff, 1933). This novelty-triggered boost of memory is absent in patients with MTL-lesion (Kishiyama et al., 2004), and such lesions similarly abolish novelty-sensitive event-related potentials responses (Knight, 1996). In contrast, in subjects with intact MTL function, electrophysiological signals such as the P300 depend on the novelty of the stimulus, and their occurrence or absence is predictive of future memory (Fabiani & Donchin, 1995; Otten & Donchin, 2000; Paller, Kutas, & Mayes, 1987). Thus, stimulus novelty triggers distinct processes in the brain. In our data, items later correctly remembered (i.e., items for which a memory is successfully formed) showed a stronger pupil constriction during memorization than later forgotten items. This paralleled the stronger constriction for novel as compared to familiar items during retrieval. Taken together, these results are consistent with the “novelty-triggers encoding” hypothesis. Under this hypothesis, the items later forgotten evoke a weaker constriction during memori-

zation because they are not detected as novel and therefore not properly encoded. In turn, the later remembered items evoke a stronger pupil constriction during encoding as they are correctly perceived as novel, therefore encoded and thus correctly remembered later. A multitude of factors, such as attention, motivation, arousal, interest in the specific stimuli shown, and previous knowledge about the items presented, contributes to memory and the likelihood that a lasting memory will be encoded. All these will influence how subjectively novel a stimulus appears to a subject engaged in a learning task.

The neural origin of the novelty signals that influence the pupil are at present unknown. While there are common neural inputs to the brain areas involved in such learning (such as the MTL) as well as the areas modulating pupil size, such a direct relationship remains to be shown. We suggest that our task would be a candidate for such an experiment, either combining pupillometry with imaging or electrophysiology, for which the present data provide clear initial hypotheses (see below). It should be noted, however, that due to the comparably large intersubject variability, such an experiment needs to be performed in a within-subject design, ideally with simultaneous recordings of the neural and pupillometric data, which is prohibitive of performing a direct comparison between the present data and, for example, the data of Rutishauser et al. (2010).

Whereas relations between working memory load and pupil dilation have long been established (Gardner, Beltramo, & Krinsky, 1975; Gardner, Philp, & Radacy, 1978; Granholm et al., 1996; Kahneman & Beatty, 1966), the relationship between pupil dilation and declarative memory for novel natural scenes has remained unclear. We used a 6-point confidence scale during retrieval combined with analysis of ROC slopes to verify that our observers had established declarative memories. The same method for confidence judgments has been used by others to distinguish between different models of memory according to whether an effect is observed at certain or all levels of confidence (for details, see Wixted, 2007). While evaluating specific memory models is beyond the scope of the present study, we intentionally designed our task virtually identical to tasks that have been used previously in a number of behavioral, functional imaging and intracranial recording studies on memory (Rutishauser et al., 2010; Wais et al., 2006; Wixted, 2007). Our results show that the pupil size contains information related to aspects of declarative memory formation and retrieval, thus offering a metric of potential interest for future studies of similar tasks. Our findings indicate that the stronger constriction that follows novel stimuli can be utilized to assess components of both memory formation and retrieval. We deliberately used natural scenes of varying luminance to demonstrate that this effect

can be observed in relatively realistic situations. This is in contrast to some of the few previous studies that addressed familiarity effects in the pupil with words (Heaver & Hutton, 2011; Otero et al., 2011; Vö et al., 2008). Observers were asked to study a list of words and were later asked to indicate for another list of words whether the word was present on the previously studied list (Heaver & Hutton, 2011; Otero et al., 2011; Vö et al., 2008). Importantly, none of the words were novel as such. Hence, these studies did not address the formation of memory for novel items and found no significant relationship between pupil dilation and later memory performance (Vö et al., 2008). In contrast, we demonstrate that pupil dilation during memory formation for novel natural scenes is predictive of later memory strength, i.e., it exhibits a difference-due-to-memory effect (Paller & Wagner, 2002) similar to what has previously been found using either invasive recordings or imaging (Johnson et al., 2009; Rutishauser et al., 2010).

A previous study (Kafkas & Montaldi, 2011) similarly found that the absolute size of the pupil during *incidental* encoding is more constricted for later remembered scenes than for later forgotten scenes. Our study confirms this finding for both incidental and effortful memory encoding, the latter of which the previous study did not address. Unlike the present study, Kafkas and Montaldi (2011) based their conclusion on the *absolute* size of the pupil. Absolute measures are prone to drifts and deflections remaining from previous trials. Although these authors confirmed that this could not explain their results, we consider the pupil slope analyzed here to be a more robust measure that is insensitive to baseline shifts (see Methods). Nevertheless, we also confirmed that the pupil size at image onset (prestimulus baseline) during memorization did not predict memory formation in our task.

Note that in our tasks, stimuli were seen only once by observers before the memory test, indicating that the familiar-novel pupil difference can develop rapidly after single-trial exposure. This suggests that the pupil response is a valuable tool to assess memory formation. We consider it likely that simultaneous recordings of pupil diameter with other measures, such as invasive electrophysiology or brain imaging, will provide a powerful combination of signals that operate on different spatial and temporal scales and may thus enable deeper insights into the neural mechanisms of memory formation.

Hypotheses for a neural explanation of the pupil-novelty relation

Pupil size is controlled by both the sympathetic and parasympathetic pathways and modulated by many

cognitive processes (e.g., Bradshaw, 1967; Einhäuser, Koch, & Carter, 2010; Einhäuser, Stout, Koch, & Carter, 2008; Friedman, Hakerem, Sutton, & Fleiss, 1973; Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Granholm et al., 1996; Hess & Polt, 1964; Jepma & Nieuwenhuis, 2011; Kahneman & Beatty, 1966; Naber, Frässle, & Einhäuser, 2011; Raisig, Welke, Hagendorf, & van der Meer, 2010; Simpson & Hale, 1969; Yoss et al., 1970). Pupil dilation via the sympathetic pathway by cognitive processes is thought to be carried mainly by noradrenergic (NA) projections originating in the locus coeruleus (LC; Berridge & Foote, 1991; Jepma et al., 2011; Koss, 1986; Nieuwenhuis, de Geus, & Aston-Jones, 2011; Phillips, Szabadi, & Bradshaw, 2000; Rajkowski, Kubiak, & Aston-Jones, 1994; Samuels & Szabadi, 2008; Sterpenich et al., 2006; Vankov, Hervé-Minvielle, & Sara, 1995), although direct electrophysiological evidence in humans is still lacking. Interestingly, LC neurons respond vigorously to novel objects (Vankov et al., 1995), which in humans would result in more dilation, rather than in the stronger constriction observed here. Only rarely, however, has NA been related to memory processes in humans, and if so, in particular in the context of processing emotional items (Harrison, Singer, Rotshtein, Dolan, & Critchley, 2006; Sterpenich et al., 2006). From animal studies on the effect of stress hormones, such as adrenaline, on memory modulation there is, however, strong evidence that NA projections, in particular those to the basolateral nucleus of the amygdala, play an important role in memory consolidation (McGaugh & Roozendaal, 2008, for review). In addition, retrieval of weak memories also benefits from increased NA levels, whether they are induced pharmacologically, by LC stimulation or by increasing arousal through contextual cues (Sara, 2009, for review).

It is likely that cognitive factors reflected in the LC-NA signal contribute to some of the pupil size effects seen here. In particular, the increased prestimulus baseline during retrieval for correct responses is possibly related to an increased level of alertness or arousal prior to stimulus onset. Importantly, however, this effect is absent for the pupil slope, rendering it unlikely that the prestimulus baseline confounds the constriction during stimulus presentation. Moreover, prestimulus baseline effects are not seen during memorization. Together with the inversion of the effects on constriction between memorization and retrieval render it unlikely that factors other than familiarity/novelty contribute strongly to the observed differences in the present paradigm. It should be noted, however, that the present task (memorization of whole scenes) is comparably simple (as evidenced by the high performance), as is the recognition of the images as such. In addition, declarative memory for pictures seems to have very little intrinsic capacity limitations

(Standing, 1973; Standing, Conezio, & Haber, 1970), which is distinct from working or short-term memory. To address the question, how other cognitive factors that are mediated through LC-NA—such as, for example, task (dis)engagement or an exploitation/exploration tradeoff (cf. Gilzenrat et al., 2010; Jepma & Nieuwenhuis, 2011), sensory uncertainty, attentional and capacity limitations, emotional processes, arousal, and alertness—relate to our results on novelty/familiarity would require a more challenging memorization task. While clearly beyond the scope of the present study, we consider linking the different factors acting on the pupil through the LC-NA system to the results presented here one of the most exciting questions in pupillometric research.

In contrast to NA, acetylcholine (ACh), which leads to pupil constriction through the parasympathetic nervous system (Hasselmo & Giocomo, 2006) and is best known for mediating the pupil light reflex (Loewenfeld & Lowenstein, 1993), seems to play an important role in memory processes: (a) ACh modulates memory encoding in rats (Hasselmo, 2006; Hasselmo & Giocomo, 2006; Warburton et al., 2003), which has also been shown in humans (Kukolja, Thiel, & Fink, 2009; Silver, Shenhav, & D’Esposito, 2008) with pharmacological manipulations; (b) Patients with Alzheimer’s disease show attenuated pupillary constrictions (Fountoulakis, St Kaprinis, & Fotiou, 2004; Prettyman, Bitsios, & Szabadi, 1997), and reduced levels of ACh (Francis, Palmer, Snape, & Wilcock, 1999), which is similarly shown in the relation between average constriction and memory performance found for healthy individuals in the present study (cf. Figure 7); (c) ACh release has been related to neural plasticity (Bakin & Weinberger, 1996; Kilgard & Merzenich, 1998) and memory consolidation during rapid eye movement (REM) sleep (Power, 2004). The complex interplay between dilation and constriction renders it difficult to conclude from our data alone, whether “sympathetic dilation” or “parasympathetic constriction” is at the foundation of the effects we describe. Nonetheless, our data together with the aforementioned previous studies allows us to hypothesize that a cholinergic novelty signal underlies both the observed pupil constriction and participates in memory formation. Irrespective of the neural origin, however, the finding that pupil size predicts as to whether an item will later be remembered provides an immediately available signal that can potentially be used for online monitoring of memory formation. Such online monitoring potentially paves the way for interfering with memory formation, which then—in combination with imaging and electrophysiology—may provide a useful tool both for studying and eventually improving declarative memory formation.

Keywords: memory, natural scene, pupil

Acknowledgments

We thank Steffen Klingenhoefer for discussion. This work was supported by the German Research Foundation (DFG) through Research Training Group 885 “NeuroAct” – (MN), grants EI852/1 and EI852/3 (WE), and by the Max Planck Society (UR).

* Authors MN and SF contributed equally to this article.

Commercial relationships: none.

Corresponding author: Wolfgang Einhäuser.

Email: wet@physik.uni-marburg.de.

Address: Philipps-University Marburg, AG Neurophysik, Marburg, Germany.

References

- Bakin, J. S., & Weinberger, N. M. (1996). Induction of a physiological memory in the cerebral cortex by stimulation of the nucleus basalis. *Proceedings of the National Academy of Sciences USA*, 93(20), 11219–11224.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57, 289–300.
- Berridge, C. W., & Foote, S. L. (1991). Effects of locus coeruleus activation on electroencephalographic activity in neocortex and hippocampus. *Journal of Neuroscience*, 11(10), 3135–3145.
- Bitsios, P., Szabadi, E., & Bradshaw, C. M. (2002). Relationship of the ‘fear-inhibited light reflex’ to the level of state/trait anxiety in healthy subjects. *International Journal of Psychophysiology*, 43(2), 177–184.
- Bradshaw, J. (1967). Pupil size as a measure of arousal during information processing. *Nature*, 216(5114), 515–516.
- Brainard, D. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436.
- Chadwick, M., Hassabis, D., Weiskopf, N., & Maguire, E. (2010). Decoding individual episodic memory traces in the human hippocampus. *Current Biology*, 20(6), 544–547.
- Charney, D. S., Scheier, I. H., & Redmond, D. E. (1995). Noradrenergic neural substrates for anxiety and fear. In F. E. Bloom & D. J. Kupfer (Eds.), *Psychopharmacology: The fourth generation of progress* (pp. 387–396). New York: Raven Press.
- Cornelissen, F., Peters, E., & Palmer, J. (2002). The

- Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavioral Research Methods, Instruments, & Computers*, 34(4), 613–617.
- Daniels, L., Hock, H. S., & Huisman, A. (2009). The effect of spatial attention on pupil dynamics. *Perception*, 38, 16.
- Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences USA*, 93(24), 13494–13499.
- Einhäuser, W., Koch, C., & Carter, O. (2010). Pupil dilation betrays the timing of decisions. *Frontiers in Human Neuroscience*, 4, 18.
- Einhäuser, W., Stout, J., Koch, C., & Carter, O. (2008). Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. *Proceedings of the National Academy of Sciences USA*, 105(5), 1704–1709.
- Fabiani, M., & Donchin, E. (1995). Encoding processes and memory organization: A model of the von Restorff effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21(1), 224–240.
- Fountoulakis, K. N., St Kaprinis, G., & Fotiou, F. (2004). Is there a role for pupillometry in the diagnostic approach of Alzheimer's disease? A review of the data. *Journal of the American Geriatrics Society*, 52(1), 166–168.
- Francis, P. T., Palmer, A. M., Snape, M., & Wilcock, G. K. (1999). The cholinergic hypothesis of Alzheimer's disease: a review of progress. *Journal of Neurology, Neurosurgery & Psychiatry*, 66(2), 137–147.
- 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 & Clinical Neurophysiology*, 34(5), 475–484.
- Gardner, R. M., Beltramo, J. S., & Krinsky, R. (1975). Pupillary changes during encoding, storage, and retrieval of information. *Perceptual & Motor Skills*, 41(3), 951–955.
- Gardner, R. M., Philp, P., & Radacy, S. (1978). Pupillary changes during recall in children. *Journal of Experimental Child Psychology*, 25(1), 168–172.
- Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (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.
- Gonsalves, B., Kahn, I., Curran, T., Norman, K., & Wagner, A. (2005). Memory strength and repetition suppression: Multimodal imaging of medial temporal cortical contributions to recognition. *Neuron*, 47(5), 751–761.
- Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, 33(4), 457–461.
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, 10(1), 14–23.
- Harrison, N., Singer, T., Rotshtein, P., Dolan, R., & Critchley, H. (2006). Pupillary contagion: Central mechanisms engaged in sadness processing. *Social Cognitive & Affective Neuroscience*, 1(1), 5–17.
- Hassabis, D., Chu, C., Rees, G., Weiskopf, N., Molyneux, P., & Maguire, E. (2009). Decoding neuronal ensembles in the human hippocampus. *Current Biology*, 19(7), 546–554.
- Hasselmo, M. E. (2006). The role of acetylcholine in learning and memory. *Current Opinion in Neurobiology*, 16(6), 710–715.
- Hasselmo, M. E., & Giocomo, L. M. (2006). Cholinergic modulation of cortical function. *Journal of Molecular Neuroscience*, 30(1–2), 133–135.
- Heaver, B., & Hutton, S. B. (2011). Keeping an eye on the truth? Pupil size changes associated with recognition memory. *Memory*, 19(4), 398–405.
- Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, 143(3611), 1190–1192.
- Jepma, M., Deinum, J., Asplund, C. L., Rombouts, S. A., Tamsma, J. T., Tjeerdema, N., et al. (2011). Neurocognitive function in dopamine- β -hydroxylase deficiency. *Neuropsychopharmacology*, 36(8), 1608–1619.
- Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration-exploitation trade-off: Evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience*, 23(7), 1587–1596.
- Johnson, J., McDuff, S., Rugg, M., & Norman, K. (2009). Recollection, familiarity, and cortical reinstatement: A multivoxel pattern analysis. *Neuron*, 63(5), 697–708.
- Kafkas, A., & Montaldi, D. (2011). Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *Quarterly Journal of Experimental Psychology (Hove)*, 64(10), 1971–1989.

- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(756), 1583–1585.
- Karatekin, C. (2004). Development of attentional allocation in the dual task paradigm. *International Journal of Psychophysiology*, *52*(1), 7–21.
- Karatekin, C., Couperus, J. W., & Marcus, D. J. (2004). Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses. *Psychophysiology*, *41*(2), 175–185.
- Kilgard, M. P., & Merzenich, M. M. (1998). Cortical map reorganization enabled by nucleus basalis activity. *Science*, *279*(5357), 1714–1718.
- Kishiyama, M. M., Yonelinas, A. P., & Lazzara, M. M. (2004). The von Restorff effect in amnesia: The contribution of the hippocampal system to novelty-related memory enhancements. *Journal of Cognitive Neuroscience*, *16*(1), 15–23.
- Knight, R. (1996). Contribution of human hippocampal region to novelty detection. *Nature*, *383*(6597), 256–259.
- Koss, M. C. (1986). Pupillary dilation as an index of central nervous system alpha 2-adrenoceptor activation. *Journal of Pharmacological Methods*, *15*(1), 1–19.
- Kukulja, J., Thiel, C. M., & Fink, G. R. (2009). Cholinergic stimulation enhances neural activity associated with encoding but reduces neural activity associated with retrieval in humans. *Journal of Neuroscience*, *29*(25), 8119–8128.
- Kumaran, D., & Maguire, E. (2009). Novelty signals: A window into hippocampal information processing. *Trends in Cognitive Sciences*, *13*(2), 47–54.
- Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences USA*, *99*(14), 9596–9601.
- Lisman, J. E., & Grace, A. A. (2005). The hippocampal-VTA loop: Controlling the entry of information into long-term memory. *Neuron*, *46*(5), 703–713.
- Loewenfeld, I., & Lowenstein, O. (1993). *The pupil: Anatomy, physiology, and clinical applications*. Detroit: Wayne State University Press.
- Macmillan, N., & Creelman, C. (2005). *Detection theory* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Maw, N. N., & Pomplun, M. (2004). Studying human face recognition with the gaze-contingent window technique. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Meeting of the Cognitive Science Society*. (pp. 927–932). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- McDuff, S., Frankel, H., & Norman, K. (2009). Multivoxel pattern analysis reveals increased memory targeting and reduced use of retrieved details during single-agenda source monitoring. *Journal of Neuroscience*, *29*(2), 508–516.
- McGaugh, J. L., & Roozendaal, B. (2008). Drug enhancement of memory consolidation: Historical perspective and neurobiological implications. *Psychopharmacology (Berlin)*, *202*(1–3), 3–14.
- Montaldi, D., Spencer, T., Roberts, N., & Mayes, A. (2006). The neural system that mediates familiarity memory. *Hippocampus*, *16*(5), 504–520.
- Naber, M., Frässle, S., & Einhäuser, W. (2011). Perceptual rivalry: Reflexes reveal the gradual nature of visual awareness. *PLoS One*, *6*(6), e20910.
- Nagai, M., Wada, M., & Sunaga, N. (2002). Trait anxiety affects the pupillary light reflex in college students. *Neuroscience Letters*, *328*(1), 68–70.
- Nieuwenhuis, S., de Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*, *48*, 162–175.
- Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology*, *4*, 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.
- Paller, K. A., Kutas, M., & Mayes, A. R. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography & Clinical Neurophysiology*, *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.
- Phillips, M. A., Szabadi, E., & Bradshaw, C. M. (2000). Comparison of the effects of clonidine and yohimbine on spontaneous pupillary fluctuations in healthy human volunteers. *Psychopharmacology (Berlin)*, *150*(1), 85–89.
- Polyn, S., Natu, V., Cohen, J., & Norman, K. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, *310*(5756), 1963–1966.
- Power, A. E. (2004). Slow-wave sleep, acetylcholine, and memory consolidation. *Proceedings of the*

- National Academy of Sciences USA*, 101(7), 1795–1796.
- Prettyman, R., Bitsios, P., & Szabadi, E. (1997). Altered pupillary size and darkness and light reflexes in Alzheimer's disease. *Journal of Neurology, Neurosurgery, & Psychiatry*, 62(6), 665–668.
- Raisig, S., Welke, T., Hagedorf, H., & van der Meer, E. (2010). I spy with my little eye: Detection of temporal violations in event sequences and the pupillary response. *International Journal of Psychophysiology*, 76(1), 1–8.
- Rajkowski, J., Kubiak, P., & Aston-Jones, G. (1994). Locus coeruleus activity in monkey: Phasic and tonic changes are associated with altered vigilance. *Brain Research Bulletin*, 35(5–6), 607–616.
- Ranganath, C., & Blumenfeld, R. (2005). Doubts about double dissociations between short- and long-term memory. *Trends in Cognitive Sciences*, 9(8), 374–380.
- Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nature Review Neuroscience*, 4(3), 193–202.
- Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, 99, 518–535.
- Rissman, J., Greely, H., & Wagner, A. (2010). Detecting individual memories through the neural decoding of memory states and past experience. *Proceedings of the National Academy of Sciences USA*, 107(21), 9849–9854.
- Rutishauser, U., Mamelak, A., & Schuman, E. (2006). Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. *Neuron*, 49(6), 805–813.
- Rutishauser, U., Ross, I. B., Mamelak, A. N., & Schuman, E. M. (2010). Human memory strength is predicted by theta-frequency phase-locking of single neurons. *Nature*, 464(7290), 903–907.
- Rutishauser, U., Schuman, E., & Mamelak, A. (2008). Activity of human hippocampal and amygdala neurons during retrieval of declarative memories. *Proceedings of the National Academy of Sciences USA*, 105(1), 329–334.
- Samuels, E. R., & Szabadi, E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: Its roles in the regulation of arousal and autonomic function part II: Physiological and pharmacological manipulations and pathological alterations of locus coeruleus activity in humans. *Current Neuropharmacology*, 6(3), 254–285.
- Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, 10(3), 211–223.
- Silver, M. A., Shenhav, A., & D'Esposito, M. (2008). Cholinergic enhancement reduces spatial spread of visual responses in human early visual cortex. *Neuron*, 60(5), 904–914.
- Simpson, H. M., & Hale, S. M. (1969). Pupillary changes during a decision-making task. *Perceptual & Motor Skills*, 29(2), 495–498.
- Simpson, H. M., & Molloy, F. M. (1971). Effects of audience anxiety on pupil size. *Psychophysiology*, 8(4), 491–496.
- Sokolov, E. N. (1963). Higher nervous functions: The orienting reflex. *Annual Review of Physiology*, 25, 545–580.
- Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, 27, 279–306.
- Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: A new perspective. *Nature Review Neuroscience*, 8(11), 872–883.
- Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, 25(2), 207–222.
- Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 19, 73–74.
- Steinhauer, S. R., Boller, F., Zubin, J., & Pearlman, S. (1983). Pupillary dilation to emotional visual stimuli revisited. *Psychophysiology*, 20, 472.
- Sterpenich, V., D'Argembeau, A., Desseilles, M., Balteau, E., Albouy, G., Vandewalle, G., et al. (2006). The locus coeruleus is involved in the successful retrieval of emotional memories in humans. *Journal of Neuroscience*, 26(28), 7416–7423.
- Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, 53, 1–25.
- Vankov, A., Hervé-Minvielle, A., & Sara, S. J. (1995). Response to novelty and its rapid habituation in locus coeruleus neurons of the freely exploring rat. *European Journal of Neuroscience*, 7(6), 1180–1187.
- Viskontas, I., Knowlton, B., Steinmetz, P., & Fried, I. (2006). Differences in mnemonic processing by neurons in the human hippocampus and parahippocampal regions. *Journal of Cognitive Neuroscience*, 18(10), 1654–1662.
- Võ, M. L., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., et al. (2008). The

- coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophysiology*, 45(1), 130–140.
- von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung*, 18, 299–342.
- Wais, P. E., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2006). The hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron*, 49, 459–466.
- Warburton, E. C., Koder, T., Cho, K., Massey, P. V., Duguid, G., Barker, G. R., et al. (2003). Cholinergic neurotransmission is essential for perirhinal cortical plasticity and recognition memory. *Neuron*, 38(6), 987–996.
- Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, 114(1), 152–176.
- Xiang, J., & Brown, M. (1998). Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. *Neuropharmacology*, 37(4–5), 657–676.
- Yonelinas, A., Otten, L., Shaw, K., & Rugg, M. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience*, 25(11), 3002–3008.
- Yoss, R. E., Moyer, N. J., & Hollenhorst, R. W. (1970). Pupil size and spontaneous pupillary waves associated with alertness, drowsiness, and sleep. *Neurology*, 20(6), 545–554.