Behavioral/Cognitive

Pupil-Linked Arousal Responds to Unconscious Surprisal

Andrea Alamia,[1,2] Rufin VanRullen,[2] Emanuele Pasqualotto,[1] André Mouraux,[1] and Alexandre Zenon[1,3]
[1]Institute of Neuroscience, Université Catholique de Louvain, Brussels 1200, Belgium, [2]CerCo, CNRS, Université de Toulouse, Toulouse 31052, France, and [3]Institut de Neurosciences Cognitives et Intégratives d'Aquitaine, 33076 Bordeaux, France

Pupil size under constant illumination reflects brain arousal state, and dilates in response to novel information, or surprisal. Whether this response can be observed regardless of conscious perception is still unknown. In the present study, male and female adult humans performed an implicit learning task across a series of three experiments. We measured pupil and brain-evoked potentials to stimuli that violated transition statistics but were not relevant to the task. We found that pupil size dilated following these surprising events, in the absence of awareness of transition statistics, and only when attention was allocated to the stimulus. These pupil responses correlated with central potentials, evoking an anterior cingulate origin. Arousal response to surprisal outside the scope of conscious perception points to the fundamental relationship between arousal and information processing and indicates that pupil size can be used to track the progression of implicit learning.

Key words: **arousal; ERPs; implicit learning; prediction error; pupil size; statistical learning**

Significance Statement

Pupil size dilates following increase in mental effort, surprise, or more generally global arousal. However, whether this response arises as a conscious response or reflects a more fundamental mechanism outside the scrutiny of awareness is still unknown. Here, we demonstrate that unexpected changes in the environment, even when processed unconsciously and without being relevant to the task, lead to an increase in arousal levels as reflected by the pupillary response. Further, we show that the concurrent electrophysiological response shares similarities with mismatch negativity, suggesting the involvement of anterior cingulate cortex. All in all, our results establish novel insights about the mechanisms driving global arousal levels, and it provides new possibilities for reliably measuring unconscious processes.

Introduction

The main function of the eye pupil is to adjust the quantity of light reaching the retina by dilating or constricting in response to environmental luminance changes (De Groot and Gebhard, 1952). However, pupil size is also affected by global brain arousal state (Bradley et al., 2008; Reimer et al., 2016), which is modulated by a wide range of cognitive factors (Hess and Polt, 1964; Beatty and Lucero-Wagoner, 2000; Bradley et al., 2008; Mathot et al., 2014). The fundamental, causal link between cognition and arousal remains unknown but authors have hypothesized that uncertainty, and its reduction through brain processing, may be the common denominator of arousal responses to cognition (Yu and Dayan, 2005; Preuschoff et al., 2011; Zénon et al., 2019). In agreement with this view, pupil size has been shown to respond to self-information, or surprisal, which can be defined as the negative log probability of an event, i.e., how unexpected an observation is (Friedman et al., 1973; Raisig et al., 2010; Preuschoff et al., 2011; Nassar et al., 2012; O'Reilly et al., 2013; Damsma and van Rijn, 2017). Nonetheless, these results were obtained in the context of behavioral tasks, in which the surprising observation indicated the need for behavioral adaptation to new contingencies, and possibly emotional surprise response (Meyer et al., 1997). Although a few studies have reported pupil dilation to surprising stimuli in the absence of behavioral responses (Gomes et al., 2015; Liao et al., 2016; Zekveld et al., 2018), these findings were obtained in contexts in which surprising events were conscious.

Consequently, it remains unknown whether the eye pupil responds to unconscious violations of predictions, or whether it relies on conscious processes. Such unconscious predictions occur, for instance, in the context of statistical learning, the process through which we learn the statistical regularities of the environment (Kim et al., 2009; Turk-Browne et al., 2009). This process has been reported for the first time in the auditory domain, showing that 8-months babies were able to learn differences in statistical relationship between non-meaningful syllables (Saffran et

Received Nov. 29, 2018; revised April 26, 2019; accepted April 27, 2019.
Author contributions: A.A. and A.Z. designed research; A.A. and E.P. performed research; A.A., R.V., E.P., A.M., and A.Z. analyzed data; A.A. and A.Z. wrote the paper.
This work was supported by Grants from the Fondation Médicale Reine Elisabeth, the Fonds de la Recherche Scientifique, and the Fondation Louvain and IdEx Bordeaux.
The authors declare no competing financial interests.
Correspondence should be addressed to Andrea Alamia at andrea.alamia@cnrs.fr.
https://doi.org/10.1523/JNEUROSCI.3010-18.2019

al., 1996); since then, statistical learning has been demonstrated in many other studies, for example in the context of language acquisition (Saffran et al., 1999; Arciuli and Torkildsen, 2012), but also in visual processing (Turk-browne et al., 2005, 2008; Alamia and Zénon, 2016). Evidence for pupillary response to violations of expectations in statistical learning would also make pupil size a compelling physiological marker to track learning progression.

Here, we tested whether pupil size reacts to the violation of statistical regularities when these are not consciously perceived as such by participants and, most importantly, the violations are not task-relevant. In a series of three experiments, we replicated the results investigating the role of attention (Experiment 1), the level of rules' awareness (Experiment 2) and finally the electrophysiological correlates of the pupillary response (Experiment 3).

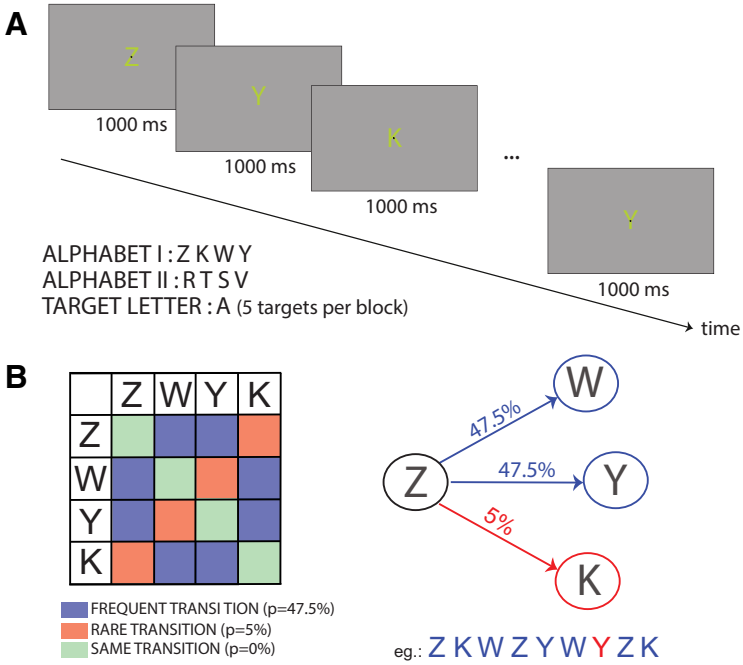

Figure 1. Experimental design. ***A***, Each trial was composed of a series of displays containing a letter over-imposed on a fixation point. The letters changed every second. In the first experiment, participants underwent 2 sessions, with a different alphabet in each session (Alphabets I or II). Experiments 2 and 3 used only Alphabet I. ***B***, Color-coded transition matrix for rare (red), frequent (blue), and same transitions (green). Notice that the same transition never occurred during the experiment. Right, A schematic example of the Markovian process driving the letter sequence. Each letter could transition to three other letters, two of which being frequent (47.5% chance) and one being rare (5% chance). Bottom, An example of sequence is provided, in which the rare transition is shown in red.

Materials and Methods

Participants. Fifty-two healthy participants (Experiment 1: 14 participants, 4 females, mean age = 28.64 years, SD = 5.06; Experiment 2: 18 participants, 12 females, mean age = 24.77 years, SD = 2.88; Experiment 3: 20 participants, 11 females, mean age = 22.41 years, SD = 2.58) took part in this study. The initial sample size of 14 participants for the first experiment was chosen based on a previous pilot study, performed with the same number of participants. We used a larger sample size in Experiments 2 and 3, to ensure statistical power given that we expected negative results in the familiarity and generative tests, and to compensate for the limited number of trials in the rare condition for the electrophysiological analysis. Two participants from Experiments 3 were rejected because of technical failures in the electrophysiological recordings. All participants reported normal or corrected-to-normal vision. All experiments were performed according to the Declaration of Helsinki and were approved by the Ethics Committee of the Université Catholique de Louvain. Written informed consents were obtained from all the participants, who received monetary compensation for their participation.

Procedure. The experiment took place in a dim room, with the participants sitting in front of a 19 inch CRT screen with a 100 Hz refresh rate. The distance between the screen and the chin support was 58 cm, whereas the height of the chin support was adjusted to each participant to ensure a comfortable position. The task was implemented using the version 3.0.9 of the Psychtoolbox (Brainard, 1997) in MATLAB 7.5 (MathWorks).

In all experiments, each block began with the display of a black fixation point in the center of the screen, over-imposed on a letter (Fig. 1*A*). The letters were 1° wide and 2° high. In the first experiment, participants were pseudo-randomly assigned to one of two sets of four letters for each session (Set 1: ZYKW, Set 2: RLMQ), whereas in all the other experiment they were assigned to the first set of letters. The letter, displayed in yellow (RGB: 0.9 .9 0.7) over a gray background (RGB: 0.7 .7 0.7), changed with a constant rate of one letter per second. The stream of letters followed a probabilistic Markovian process, as reported in Figure 1*B*: two transitions had a 0.475 probability of occurring (from L1 to L2 and to L3), whereas the last one had a probability of 0.050 (L1–L4).

In the first experiment, participants performed 2 sessions: each session was composed of three blocks of 5 min each (300 transitions), and a short break of few minutes was provided between sessions. The session order and the set of letters was pseudo-randomized across subjects. In one session, participants were instructed to report the appearance of a fifth letter by clicking a mouse button (the target letter was "A"); in the other session, they were instructed to click whenever the fixation point turned from black to white for 300 ms, concurrently with the onset of a new trial (i.e., the onset of a new letter). No target letter (i.e., "A") was displayed in this session. In each block, only five targets (either the "A" or the change in the fixation point color) were displayed during the experiment such that the target events occurred each at a random point within successive epochs lasting one fifth of the total task duration. In the second and third experiments, participants performed, respectively, six and three blocks of the letter detection task, in which they were instructed to detect the letter "A", similarly to the one session of the first experiment. In all experiments participants were not informed of the presence of the rules and, during the final debriefing at the end of the experiment, none of them was able to report the nature of the rules. In the second experiment, after the end of the last block, participants were informed about the presence of probabilistic rules (without detailing them yet), and were asked to perform one familiarity and one generative task. The familiarity task consisted in providing a rate of familiarity from 1 (not familiar) to 10 (very familiar) to all the possible letters' transitions. During the generative task, participants were instructed to guess which letter would be the most likely to follow after an initial two-letter transition. In both experiments, every transition was presented seven times in random order.

Statistical analysis. All Bayesian statistical analyses were performed in JASP (Love et al., 2015; JASP Team, 2018). Ultimately, the goal of Bayesian analyses is to compute Bayes factors (BFs), which quantify the ratio between the model evidence of two alternative statistical models. If not otherwise specified, all BFs refer to the comparison between the null and the alternative hypothesis, and are reported as BF_{10}, which means that larger values indicate more preference for the alternative hypothesis. A widely accepted interpretation of BF (Kass and Raftery, 1995) is that values >3 provide substantial evidence in favor of the alternative (null when BF_{01}) hypothesis, BFs >20 indicate strong evidence and BFs >150 correspond to very strong evidence. Conversely, a BF $<1/3$ suggests lack

of effect, whereas values between 1/3 and 3 provide inconclusive evidence (Bernardo and Smith, 2001; Masson, 2011). All our analyses, including *t* test, ANOVA, and correlations return BFs and error estimates, which we interpreted accordingly.

Pupillometry. The pupil diameter was acquired with an EyeLink 1000+ eye tracker video-based system (SR Research), recording monocularly pupil size (in arbitrary units) and eye movements with a sampling frequency of 500 Hz. Before analyzing the data, pupil traces were preprocessed to remove eye-blinks, which were identified by the blink detection algorithm implemented in the EyeLink system, and replaced by linear interpolations. Furthermore, traces were downsampled to 10 Hz, to facilitate the model fitting.

The pupil response was modeled according to an autoregressive model with exogenous input (ARX) in MATLAB R2015a (MathWorks), as described by Zénon (2017). This approach disentangles the spontaneous low-frequency oscillations of the pupil from the responses because of external stimulations. Briefly, the autoregressive approach can be viewed as being composed of two parts (Zénon, 2017). The autoregressive part accounts for the variance explained by previous values (i.e., autocorrelation within the signal). The second part accounts for the exogenous inputs, which models external factors. In our study, the design matrix included the onset of the letter, the occurrence of the rare transition and the onset of the target when present. The influence of each factor in the model is regulated by its order: the higher the order, the higher the number of samples used to predict the next sample. These orders were fitted independently for each subject. The output of the model is an impulse response for each considered factor and the innovation error, which is the difference between the actual pupil data and the prediction of the model. Once the model was fitted on each participant's data, we analyzed the impulse responses to the factor modeling the onset of the rare letter. Specifically, we extracted the signed absolute maximum value (which could be either positive or negative, depending on whether the pupil is narrowing or dilating) and its latency. The data were analyzed in JASP (JASP Team, 2018).

Electrophysiology. In the last experiment we recorded EEG signals from 32 actively shielded Ag-AgCl electrodes, mounted in an elastic cap in accordance with the extended 10-20 systems (Waveguard32 cap, Advanced Neuro Technologies). The frontal AFz electrode was used as ground, and common reference average. All the impedances were kept <5 kΩ. Signals were preprocessed applying a zero-phase Butterworth bandpass filter (0.5–30 Hz). Epochs were aligned on stimulus display, starting 200 ms before onset and lasting until the onset of the next stimulus (i.e., the subsequent letter). The signal was baseline-corrected (200 ms prestimulus baseline) and an independent component analysis was used to remove eye movements after visual inspection, and before rejecting epochs whose amplitude was >50 μV.

The analyses were performed in MATLAB (MathWorks) using EEGlab toolbox (Delorme and Makeig, 2004). First, given the lack of prior assumptions, we computed a cluster-based analysis over ERPs elicited by letter onset, regardless of the conditions. We preprocessed the data by means of a high-pass filter having a cutoff frequency at 3 Hz (a different cutoff at 1 Hz leads to identical clusters of electrodes but slightly larger time window). We then used cluster-based permutation analysis (Oostenveld et al., 2011), accounting for multiple comparisons by means of Monte Carlo simulations, setting the cluster-corrected significance threshold at 0.05. We identified two broad clusters of electrodes whose activity differed from baseline: the first one spans between 34 and 146 ms, whereas the second one between 158 and 246 ms. Both time windows share the electrodes FP1, FPZ, FP2, F7, F3, FZ, F4, F8, FC5, FC1, FC2, FC6, C3, CZ, C4, CP1, CP2, PZ, whereas the first time window also included the occipital electrodes POZ, PZ, O1, OZ, and O_2. We then computed the mean difference between rare and frequent conditions over trials, averaged across electrodes and time bins of each cluster. We tested by means of Bayesian *t* test whether the distribution mean was significantly different from zero. We also tested the correlation between averaged difference in the second cluster and the pupillary response difference between rare and frequent conditions.

Results

Pupil dilates in response to rare events

At first we aimed at investigating whether we could observe pupillary changes in response to rare events. For this purpose, in the first experiment 14 participants (4 females, mean age = 28.64 years, SD = 5.06) were instructed to stare at a stream of letters for 15 min (3 blocks of 5 min each; Fig. 1*A*; see Materials and Methods). The alphabet was composed of four letters, which changed every second as a probabilistic Markovian process: according to these rules, most of the transitions were frequent (47.5% probability) but rare transitions occurred 5% of the time (Fig. 1*B*; see Materials and Methods). Participants performed two sessions on different days. The order of the sessions was pseudorandomized across subjects. In one session, participants were instructed to focus on the letters' stream and press a button when a fifth target letter was displayed. Importantly, the onset of the target letter was independent of the rules, and occurred rarely during each block (i.e., 5 times per block, once every minute). This setting ensured that participants maintained attention on the letter stimuli.

During the whole experiment we recorded pupil traces that were subsequently analyzed by means of an autoregressive model, in which the letter onset, the target and the rare transitions were modeled as exogenous inputs (ARX model; see Materials and Methods; Zénon, 2017). This analysis returns an impulse response function (IRF) for each input, thus modeling the pupil's response to each factor. In the session in which attention allocation to the letter stream was enforced by the target detection task (Session A), we found that the signed absolute maximum value of the rare transition's IRF was significantly positive [one-sample Bayesian *t* test (JASP Team, 2018); mean = 0.274, SD = 0.303, SE = 0.081; BF = 37.93, error~3.5e-5%], thus revealing a consistent dilation of the pupil to rare events (Fig. 2*A*). As the distribution of the absolute values may not be unimodal, we also performed a nonparametric Wilcoxon rank test, confirming our result ($p = 0.035$). This finding suggests that participants track the statistics of the environment and that arousal increases in response to rare transitions.

Does attention affect pupil responses?

In the other session of Experiment 1 (Session B), no target letter was displayed, but participants were instructed to press a button whenever the fixation point turned from black to red for a few hundred milliseconds. The point of this manipulation was to divert attention away from the letter stream. The onset of the color change of the fixation point was always aligned with the letter change, and the frequency of change was identical to the frequency of target letter onsets in Session A (i.e., 5 per block, once every minute and completely orthogonal to the rules). Interestingly, the analysis revealed that when attention was allocated to the fixation dot and not on the letter stream, pupil diameter did not respond to rare transitions, in contrast to previous results (Fig. 2*B*; mean = 0.069, SD = 0.205, SE = 0.055; BF = 0.611, error~2.2e-5%; Wilcoxon rank test $p = 0.357$). We further confirmed this result by comparing directly the IRF's maximum values between Sessions A and B in a Bayesian one-way ANOVA, considering SESSION as a fixed factor (BF = 3.193, SE = 7.769e-4) and SUBJECTS as a random factor. However, because the difference in the variance of the two variables differed significantly violating the assumption of homoscedasticity (Bartlett test: $F_{(1,13)} = 6.014$, $p = 0.015$), we further tested the difference between sessions with a nonparametric test, confirming our previous results (Wilcoxon rank sum, $z = 2.41$, $p = 0.012$). Overall, these results suggest that attention plays an important role in

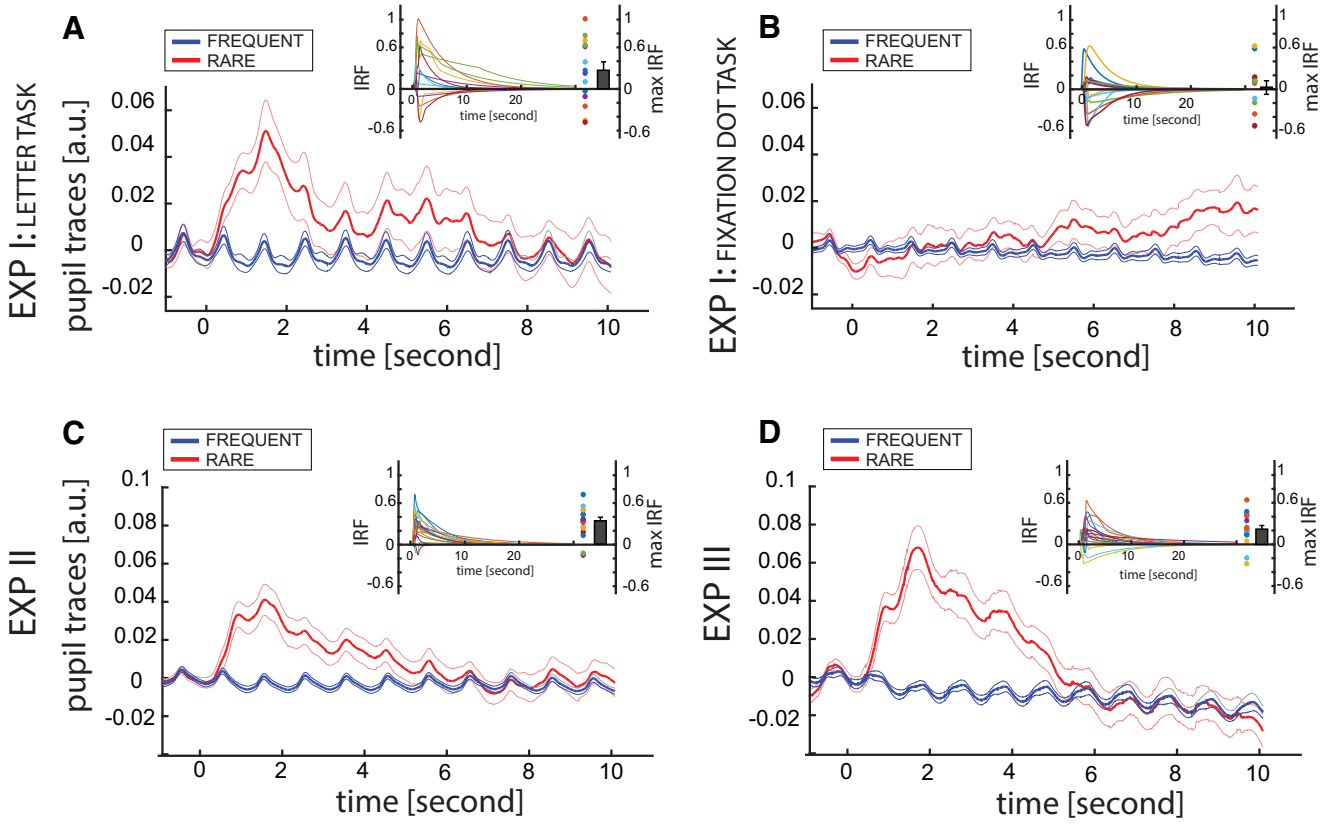

Figure 2. Pupillary traces. All the panels show pupil responses after rare (red) and frequent (blue) transitions. The origin of the *x*-axis corresponds to the onset of the stimulus, whereas the *y*-axis has arbitrary units, pupil size being normalized with respect to the baseline (1 s before onset). Baseline correction was performed for visualization purposes, but was not applied during analysis. The small graph in the top right of each panel reports the IRF of each participants computed after each rare transitions, the signed maximum value of each IRF and the mean ± SE. of the maximum values. In all panels, it is possible to observe a periodic response in both conditions at 1 Hz, because of the stimulus onset. ***A*** and ***B*** show data from the first experiment, when attention is allocated respectively to the letter or to the fixation dot. ***C***, Results from the second experiment, and (***D***) from the third one.

pupillary response to surprising events, probably by affecting the degree of statistical learning of the participants. During the debriefing at the end of the second session (i.e., Sessions A or B, depending on the pseudorandomization between subjects) none of the participants were able to explicitly report any part of the transition rules. However, to assess thoroughly the degree of awareness of the rules, we replicated the experiment on a new group of participants, adding additional tests at the end of the last block.

Level of awareness of the rules

In the second experiment, we aimed at assessing the degree of awareness of the rules, to confirm that pupil changes could be driven by unconscious events. A new pool of 18 participants (12 females, mean age = 24.77 years, SD = 2.88) performed six blocks of the target detection task (i.e., attention on the letters). As in the previous experiment, the analysis performed on the peak of the impulse response from the ARX model confirmed pupil dilation in response to rare transitions (Fig. 2*C*; mean = 0.333, SD = 0.152, SE = 0.036; BF = 1.5e12, error~2.1e-18%; Wilcoxon rank test $p = 0.0015$). Additionally, to assess participants' knowledge of the rules, they performed two further tasks immediately after the last block of the letter detection task (Destrebecqz and Peigneux, 2005; Alamia et al., 2016). The first one was a generative task, in which participants were asked to guess, according to their knowledge of the rules, which transition would follow the presentation of a pair of letters (Fig. 3*A*). The second one was a familiarity task, in which participants had to provide a rate of familiarity (1–10, the higher the more familiar the transition) to all of the possible letters' transitions (Fig. 3*B*). The generative task quantifies the capacity to use learned statistics to generate new samples and performance in this task depends on both implicit and explicit processes (Destrebecqz and Peigneux, 2005). The familiarity task evaluates the subjective recognition memory of transition statistics. At the end of the experiment, participants were verbally asked (1) whether they noticed any regularity in the stream of letters, and (2) whether they noticed some recurrent letters' series or rules driving the transitions. All participants failed to report any rule or pattern during the debriefing at the end of the experiment. However, the results from the generative task revealed some degree of knowledge of the rules. Comparison of choice probabilities for all possible choices (including the one that never occurred during the task, i.e., the double presentation of the same letter; Fig. 3*C*, green) showed robust difference between conditions (one-way Bayesian ANOVA, BF = 3.8e14, error~5.9e-6%). Difference between the same-letter and the other conditions was very significant (Bayesian paired *t* test, both BF $\gg$ 10e4, error $\ll$ 10e−11), whereas comparison between frequent and rare transitions was not conclusive (Bayesian paired *t* test: BF = 2.57, error~0.004%; Fig. 3*C*), showing that the main effect was driven by the decreased probability of choosing the same-letter transition. However, some participants tended to report some letters more often than others, regardless of transitions (χ^2 test: 8 of 18 subjects, $p < 0.05$). This behavior may have led to misleading, apparent preference for the frequent transitions in these participants. To ac-

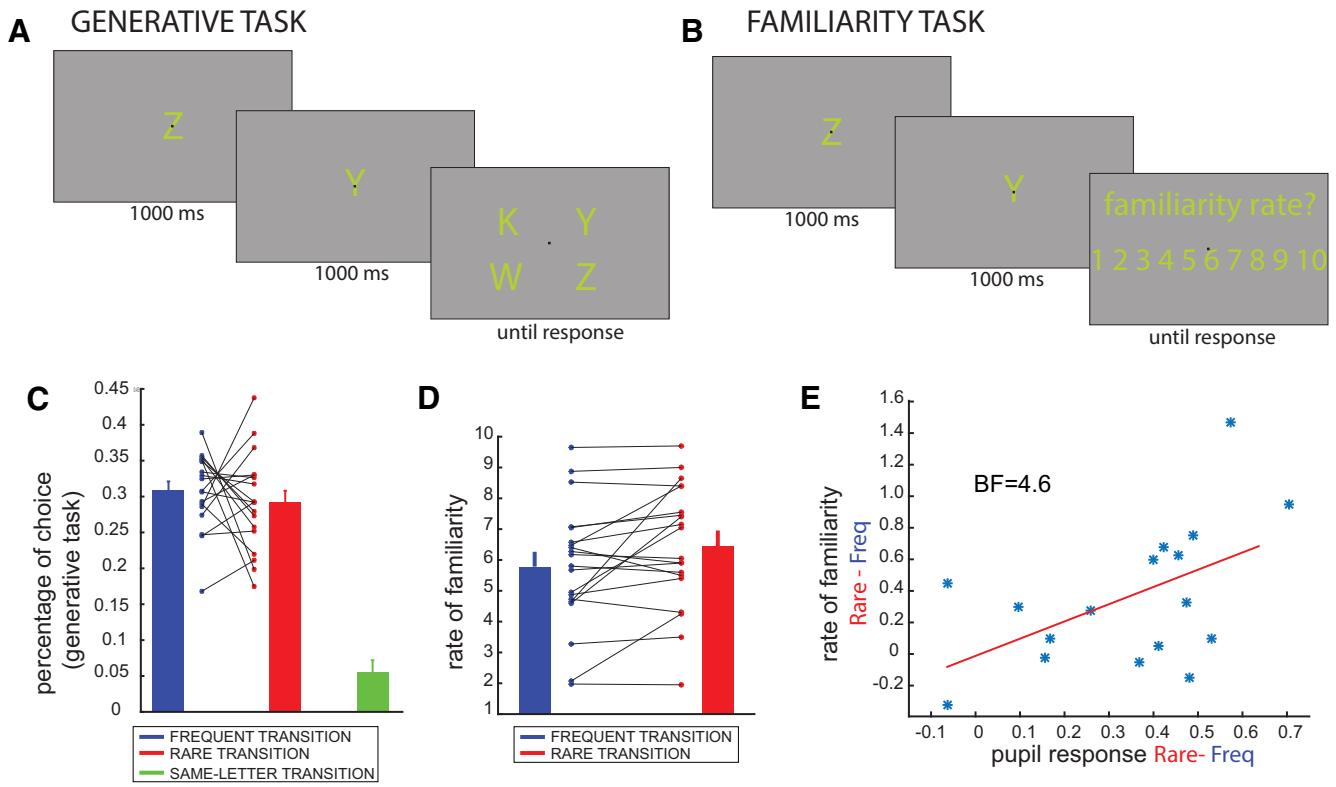

Figure 3. Generative and familiarity tests. ***A***, During a trial of the generative task, participants were asked to guess the next letter, following a transition. Letter duration and layout was the same as during the experiment. ***B***, In the familiarity task, participants judged how familiar a transition was by rating it from 1 to 10. ***C***, Results of the bias-corrected generative task: bar plots in blue/red show the mean percentage ± SE of times the participants chose the frequent/rare transition. The transitions that were chosen by the participants but never actually occurred are displayed in green (<5% of responses). The connected dots represent the percentage values for each subject in the frequent (blue) and rare (red) conditions. ***D***, Mean ± SE of the rate of familiarity reveals that participants judged rare transitions (red) as more familiar than frequent ones (blue). Each pair of connected dots represents one subject. ***E***, Positive correlation between the difference in the pupillary response between conditions (rare − frequent) and difference in the familiarity scores in each condition.

count for this confound, we rerun the analysis while expressing the dependent variable as the proportion of time each letter was reported in frequent, rare or same transitions (i.e., the probability of choosing the frequent transition was expressed as follows: $p_f = \frac{1}{4}\Sigma_{i=k,w,y,z}\, p_{i,f} = \frac{1}{4}\Sigma_{i=k,w,y,z}\frac{N_{i,f}}{N_i}$, with $N_{w,f}$ representing the number of times the letter w was chosen in the context of a frequent transition and N_w corresponding to the number of times the letter w was chosen overall). This bias correction in the data highlighted the lack of difference between frequent and rare transitions (BF in favor of null hypothesis: $BF_{01} = 3.034$, error 0.010%), thus providing more evidence for the claim that participants were not aware of the rules. Further evidence in favor of lack of awareness from the participants were revealed by the familiarity task, in which participants were asked to judge which transitions appeared more familiar to them. Surprisingly, in the familiarity task, participants judged rare transitions as more familiar than frequent ones (Bayesian paired t test: BF = 4.4, error~9.7e-4%; Fig. 3*D*). In contrast to the generative test, same-letter transitions were not assessed in the familiarity test. Moreover, in the familiarity task, we did not observe any confound because of higher/lower ratings given to specific letters, as revealed by a Bayesian ANOVA considering RATINGS as dependent variable and LETTERS and SUBJECTS respectively as fixed and random factors ($BF_{LETTERS} = 0.073$, error~3.1e-3%). A possible explanation for the larger familiarity associated with rare events is that surprisal-induced increases in arousal facilitated memorization of the rare transition (LaBar and Phelps, 1998; Mather and Sutherland, 2011). This interpretation was corroborated by a positive correlation between pupil dilation and the score difference between the two conditions (i.e., rare and frequent) in the familiarity task [Fig. 3*E*; Bayesian Pearson Correlation, $r = 0.516$, BF = 4.663, 95% credible interval: (0.083, 0.770)].

Overall, these findings show that despite some degree of learning of task structure (indicated by the arousal response) participants had no declarative awareness of the transitions (Smith and Squire, 2005) and were not able to generate or recognize frequent transitions, showing that learning occurred unconsciously.

Electrophysiological correlates

In an attempt to unveil the neural correlates of pupillary response to surprising events, we collected electrophysiological recordings (i.e., EEG) in the third and last replication of the experiment. Here, 20 participants (11 females, mean age = 22.41 years, SD = 2.58) performed three blocks of the letter detection task. As previously, we replicated the robust pupillary response to the rare transitions (Fig. 2*D*; mean = 0.193 3, SD = 0.245, SE = 0.057; BF = 19.86, error~7.3e-8%; Wilcoxon rank test $p = 0.0066$).

We first computed event related potentials (ERPs) elicited by letter transitions, i.e., regardless of the rare-frequent conditions. All ERPs were baseline-corrected by subtracting the mean of the 200 ms epoch preceding letter onset. We first determined the clusters of electrodes and the time windows-of-interest by performing a cluster-based permutation analysis, in which we compared electrodes' activity with their baseline level. This analysis

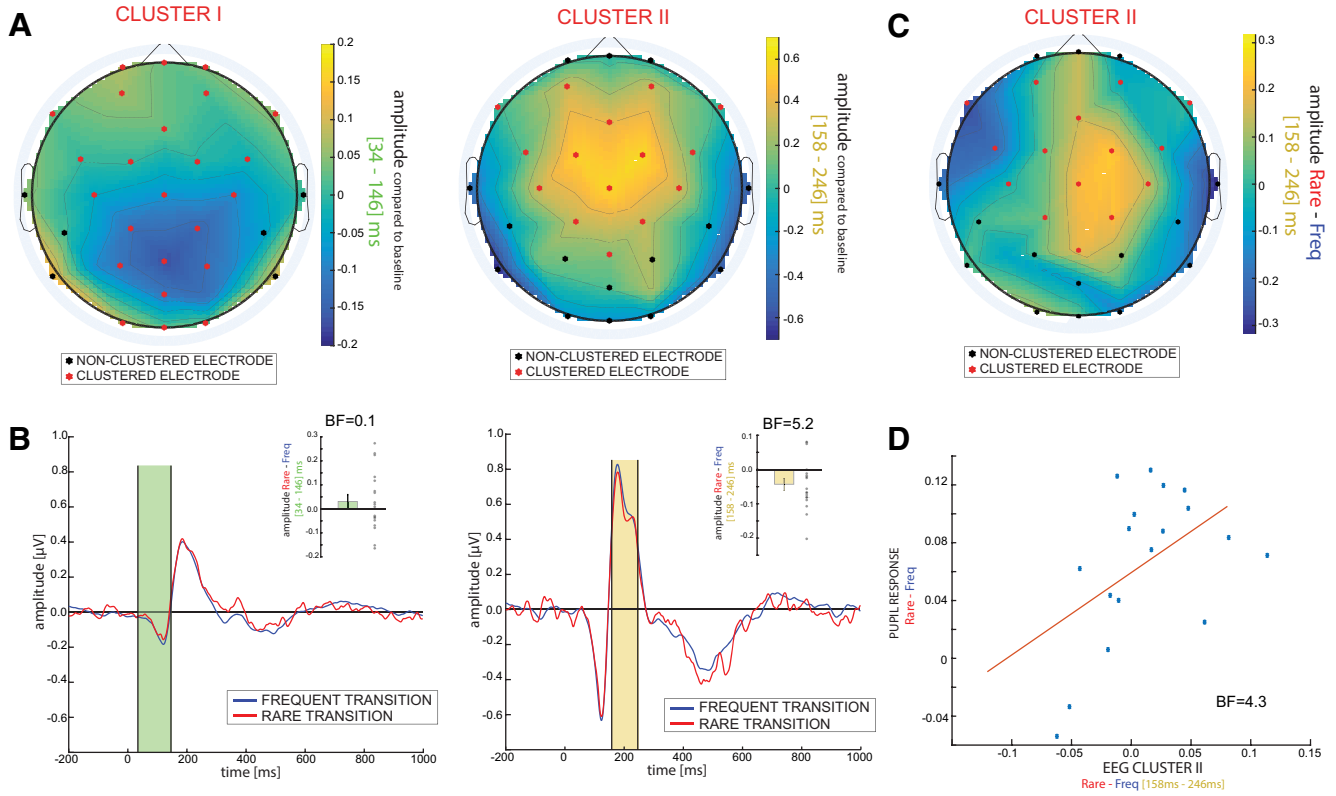

Figure 4. Electrophysiological results. ***A***, The topographies show the clusters of electrodes whose activity differed from baseline, regardless of the conditions; the left and right plots show the first and second time windows, respectively. Red/black dots represent electrodes included/excluded in/from the cluster. ***B***, Grand average of ERP traces elicited by rare (red) and frequent (blue) transitions in the first (left plot), and second cluster (right plot). The subplot shows the mean difference ± SE between rare and frequent condition in the respective time windows. Dots represent subjects. ***C***, Topography of the difference rare-frequent in the second clusters of electrodes. The effect involves mostly central regions. ***D***, Scatter plots of ERP and pupil differences (rare − frequent) in the second cluster.

revealed two broad clusters of significant activity, occurring 34–146 and 158–246 ms following letter onset (Fig. 4*A*). Notably, we found a significant difference between rare and frequent conditions in the averaged potentials of the second cluster (Bayesian one sample *t* test, BF = 5.271, err = 1.135e-4%) but not in the first cluster (BF = 0.135, err = 4.99e-4%; Fig. 4*B*). We further confirmed this result by performing a permutation-based cluster correction analysis testing the difference between frequent and rare transitions ($\alpha = 0.05$): the result of this analysis confirmed both the electrodes (as in the second cluster: frontal and central electrodes) and the time window (185–302 ms) of the Bayesian *t* tests. As shown in Figure 4*C*, the topographic distribution of amplitude difference between rare and frequent conditions in the second cluster was concentrated mostly in central electrodes. These findings are reminiscent of brain-evoked responses to deviant stimuli (e.g., statistical mismatch negativity; Koelsch et al., 2016) and suggest that these responses to surprisal originate in cingulate cortex (Fallgatter et al., 2002), a region deputed, among other functions, to monitor error detection (Carter et al., 1998; O'Connell et al., 2007) and thought to be involved in pupil dilation to cognitive processes (O'Reilly et al., 2013; Ebitz and Platt, 2015). To confirm this hypothesis, we tested the correlation between frequent-rare differences in pupil and in ERPs' responses in the second cluster, in which we observed a significant difference between the two conditions (Fig. 4*B*). As shown in Figure 4*D*, we found a robust and significant correlation [Bayesian correlation: $r = 0.496$, BF = 4.338, 95% credible interval (0.07 0.75)].

Discussion

Our results show that pupil size dilates in response to violations of expectation, even when the predictions occur unconsciously. Previous studies have already highlighted the relationship between pupil size in constant luminance and surprisal. Friedman et al. (1973) measured pupil responses to auditory sequences in which probabilities of stimulus occurrence were manipulated. They found that stimuli with low probability of occurrence, i.e., large surprisal, led to larger pupil responses than likely stimuli. More recently, pupil responses to the progressive reduction of uncertainty regarding choice outcome have been measured during various tasks. Authors found that pupil responses scaled with the divergence between prior beliefs on outcome probabilities and updated beliefs following feedback, or in other words, with the level of surprise associated with the update in outcome probabilities (Preuschoff et al., 2011; Lavín et al., 2014). Additionally, pupil size also dilates in response to events that occur at unexpected times (Kloosterman et al., 2015) or that violate prior expectations about stimulus distributions (Nassar et al., 2012; O'Reilly et al., 2013). In all these situations, however, information was directly relevant to the task, and often pertained to performance-contingent reward. Consequently, surprising information led also to policy updates (Nour et al., 2018), motivational adjustments and possibly emotional responses. The present findings that pupil responds to unconscious surprisal allow us to narrow down the range of possible causal relationships between pupil size and information processing by showing

that the mere presence of novel information, under the condition of being within the focus of attention, is sufficient to trigger pupillary dilation, independently of overt behavioral responses and conscious appraisal.

We found that, on top of pupillary responses, violations of transition expectations led to central and parietal negativity, strongly evoking classical mismatch negativity (Shelley et al., 1991; Pekkonen et al., 1995; Garrido et al., 2009; Koelsch et al., 2016), in which central negativity is observed following oddball stimuli (van Veen and Carter, 2002; Orr and Hester, 2012). Mismatch negativity has been viewed as a neural correlate of prediction error (Wacongne et al., 2012; Stefanics et al., 2014) and to originate at least partly in anterior cingulate cortex (ACC; O'Connell et al., 2007; Hyman et al., 2017). ACC is well known to encode surprise-related signals (Carter et al., 1998), to respond to task features that trigger also pupil responses (O'Reilly et al., 2013; Ebitz and Platt, 2015), and to be densely connected with locus ceruleus (Vogt et al., 2008; Joshi et al., 2016) and basal forebrain (Weible et al., 2007) thought to be among the main drivers of luminance-independent pupil responses (Aston-Jones and Cohen, 2005; Gabay et al., 2011; Eldar et al., 2013; Joshi et al., 2016; Reimer et al., 2016). Our finding that pupillary response to unconscious surprisal correlates with these brain-evoked potentials suggests either a causal relationship between ACC and pupillary responses or a common origin to these two physiological responses.

Another important finding from the present study is that rare events were reported as more familiar than frequent ones by the participants. The increased pupil response to rare events provides a clue as to the possible origin of this counter-intuitive finding: violations of expectations increased arousal, favoring the encoding of new events in memory (Nielson and Bryant, 2005; Cruciani et al., 2011; Naber et al., 2013). It is noteworthy that this preferential memorization of the rare event did not impact the responses in the generative task, indicating that unconscious and conscious processes affected generative and familiarity tasks differently (Destrebecqz and Peigneux, 2005). This suggests that the two tasks rely on different types of memory encoding, akin to the distinction between familiarity and recollection in recognition memory (Daselaar et al., 2006; Evans and Wilding, 2012). Results from the generative and familiarity tests, together with those from the verbal debriefing provide compelling evidence that the rules are learnt unconsciously.

Outside of fundamental implications, the present study also opens the way for using pupil responses as a marker of learning progression in implicit learning, which so far has relied mainly on behavioral measures, limiting the scope of learning that can be tested, and leading typically to small effect sizes (Turk-Browne et al., 2005; Barakat et al., 2013). Using this approach, we were able to confirm a previous result from the implicit literature, namely that spatial attention is required in order for implicit learning to take place (Turk-Browne et al., 2005; for review, see Chun and Turk-Browne, 2007), because pupillary dilation to surprisal occurred only when attention was allocated on the letter stream. A limitation of the present study is that we cannot disentangle the role of attention in allowing the occurrence of learning from its potential modulating effect on the pupillary response. For instance, would pupil dilation still occur when attention is not focused on the stimulus, but after a preceding learning phase? This will be an interesting venue for future research.

In conclusion, we show that an arousal response to unexpected uncertainty (Yu and Dayan, 2005) occurs regardless of awareness. This indicates the existence of a general relationship between prediction errors and arousal. Arousal is a global brain state associated with increased responsivity to stimulation (McGinley et al., 2015), larger learning rate (Nassar et al., 2012; Eldar et al., 2013), less influence of prior beliefs on inference and decisions (de Gee et al., 2014; Krishnamurthy et al., 2017; Urai et al., 2017), and exploratory behavior (Aston-Jones and Cohen, 2005; Gilzenrat et al., 2010; Jepma et al., 2010). One may therefore speculate that pupillary responses to surprisal reflect the adjustment of brain state to unexpected events that signal the need to learn more from the environment (Yu and Dayan, 2005; Parr and Friston, 2017; Payzan-LeNestour et al., 2013). These adjustments would involve extracting more information from sensory stimuli and stimulating exploration of states and policies whose knowledge remains uncertain. However, the question of whether this arousal response to uncertainty is indeed instrumental in decreasing uncertainty by assisting information processing remains to be addressed experimentally.

References

Alamia A, Zénon A (2016) Statistical regularities attract attention when task-relevant. Front Hum Neurosci 10:42.

Alamia A, Orban de Xivry JJ, San Anton E, Olivier E, Cleeremans A, Zenon A (2016) Unconscious associative learning with conscious cues. Neurosci Conscious 2016:niw016.

Arciuli J, Torkildsen JV (2012) Advancing our understanding of the link between statistical learning and language acquisition: the need for longitudinal data. Front Psychol 3:324.

Aston-Jones G, Cohen JD (2005) An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annu Rev Neurosci 28:403–450.

Barakat BK, Seitz AR, Shams L (2013) The effect of statistical learning on internal stimulus representations: predictable items are enhanced even when not predicted. Cognition 129:205–211.

Beatty J, Lucero-Wagoner B (2000) The pupillary system. In: Handbook of psychophysiology, Ed 2, pp 142–162. New York: Cambridge UP.

Bernardo JM, Smith AF (2001) Bayesian theory. Meas Sci Technol 12:221–222.

Bradley MM, Miccoli L, Escrig MA, Lang PJ (2008) The pupil as a measure of emotional arousal and autonomic activation. Psychophysiology 45:602–607.

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436.

Carter CS, Braver TS, Barch DM, Botvinick MM, Noll D, Cohen JD (1998) Anterior cingulate cortex, error detection, and the online monitoring of performance. Science 280:747–749.

Chun MM, Turk-Browne NB (2007) Interactions between attention and memory. Curr Opin Neurobiol 17:177–184.

Cruciani F, Berardi A, Cabib S, Conversi D (2011) Positive and negative emotional arousal increases duration of memory traces: common and independent mechanisms. Front Behav Neurosci 5:86.

Damsma A, van Rijn H (2017) Pupillary response indexes the metrical hierarchy of unattended rhythmic violations. Brain Cogn 111:95–103.

Daselaar SM, Fleck MS, Cabeza R (2006) Triple dissociation in the medial temporal lobes: recollection, familiarity, and novelty. J Neurophysiol 96:1902–1911.

de Gee JW, Knapen T, Donner TH (2014) Decision-related pupil dilation reflects upcoming choice and individual bias. Proc Natl Acad Sci U S A 111:E618–E625.

De Groot SG, Gebhard JW (1952) Pupil size as determined by adapting luminance. J Opt Soc Am 42:492–495.

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21.

Destrebecqz A, Peigneux P (2005) Methods for studying unconscious learning. Prog Brain Res 150:69–80.

Ebitz RB, Platt ML (2015) Neuronal activity in primate dorsal anterior cingulate cortex signals task conflict and predicts adjustments in pupil-linked arousal. Neuron 85:628–640.

Eldar E, Cohen JD, Niv Y (2013) The effects of neural gain on attention and learning. Nat Neurosci 16:1146–1153.

Evans LH, Wilding EL (2012) Recollection and familiarity make independent contributions to memory judgments. J Neurosci 32:7253–7257.

Fallgatter AJ, Bartsch AJ, Herrmann MJ (2002) Electrophysiological measurements of anterior cingulate function. J Neural Transm 109:977–988.
Friedman D, Hakerem G, Sutton S, Fleiss JL (1973) Effect of stimulus uncertainty on the pupillary dilation response and the vertex evoked potential. Electroencephalogr Clin Neurophysiol 34:475–484.
Gabay S, Pertzov Y, Henik A (2011) Orienting of attention, pupil size, and the norepinephrine system. Atten Percept Psychophys 73:123–129.
Garrido MI, Kilner JM, Stephan KE, Friston KJ (2009) The mismatch negativity: a review of underlying mechanisms. Clin Neurophysiol 120:453–463.
Gilzenrat MS, Nieuwenhuis S, Jepma M, Cohen JD (2010) Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. Cogn Affect Behav Neurosci 10:252–269.
Gomes CA, Montaldi D, Mayes A (2015) The pupil as an indicator of unconscious memory: introducing the pupil priming effect. Psychophysiology 52:754–769.
Hess EH, Polt JM (1964) Pupil size in relation to mental activity during simple problem-solving. Science 143:1190–1192.
Hyman JM, Holroyd CB, Seamans JK (2017) A novel neural prediction error found in anterior cingulate cortex ensembles. Neuron 95:447–456.e3.
JASP Team (2018) JASP (version 0.8.6.0). Amsterdam: JASP.
Jepma M, Te Beek ET, Wagenmakers EJ, van Gerven JM, Nieuwenhuis S (2010) The role of the noradrenergic system in the exploration-exploitation trade-off: a psychopharmacological study. Front Hum Neurosci 4:170.
Joshi S, Li Y, Kalwani RM, Gold JI (2016) Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. Neuron 89:221–234.
Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90:773–795.
Kim R, Seitz A, Feenstra H, Shams L (2009) Testing assumptions of statistical learning: is it long-term and implicit? Neurosci Lett 461:145–149.
Kloosterman NA, Meindertsma T, van Loon AM, Lamme VA, Bonneh YS, Donner TH (2015) Pupil size tracks perceptual content and surprise. Eur J Neurosci 41:1068–1078.
Koelsch S, Busch T, Jentschke S, Rohrmeier M (2016) Under the hood of statistical learning: a statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. Sci Rep 6:19741.
Krishnamurthy K, Nassar MR, Sarode S, Gold JI (2017) Arousal-related adjustments of perceptual biases optimize perception in dynamic environments. Nat Hum Behav 1:0107.
LaBar KS, Phelps EA (1998) Arousal-mediated memory consolidation. Psychol Sci 9:490–493.
Lavín C, San Martín R, Rosales Jubal E (2014) Pupil dilation signals uncertainty and surprise in a learning gambling task. Front Behav Neurosci 7:218.
Liao HI, Yoneya M, Kidani S, Kashino M, Furukawa S (2016) Human pupillary dilation response to deviant auditory stimuli: effects of stimulus properties and voluntary attention. Front Neurosci 10:43.
Love J, Selker R, Verhagen J, Marsman M, Gronau QF, Jamil T, Smira M, Epskamp S, Wild A, Ly A, Matzke D, Wagenmakers EJ, Morey RD, Rouder JN (2015) Software to sharpen your stats. APS Obs 28:27–29.
Masson ME (2011) A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. Behav Res Methods 43:679–690.
Mather M, Sutherland MR (2011) Arousal-biased competition in perception and memory. Perspect Psychol Sci 6:114–133.
Mathot S, Dalmaijer E, Grainger J, Van der Stigchel S (2014) The pupillary light response reflects exogenous attention and inhibition of return. J Vis 14:7 1–9.
McGinley MJ, Vinck M, Reimer J, Batista-Brito R, Zagha E, Cadwell CR, Tolias AS, Cardin JA, McCormick DA (2015) Waking state: rapid variations modulate neural and behavioral responses. Neuron 87:1143–1161.
Meyer WU, Reisenzein R, Schützwohl A (1997) Toward a process analysis of emotions: the case of surprise. Motiv Emot 21:251–274.
Naber M, Frässle S, Rutishauser U, Einhäuser W (2013) Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. J Vis 13:11 1–20.
Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, Gold JI (2012) Rational regulation of learning dynamics by pupil-linked arousal systems. Nat Neurosci 15:1040–1046.
Nielson KA, Bryant T (2005) The effects of non-contingent extrinsic and intrinsic rewards on memory consolidation. Neurobiol Learn Mem 84:42–48.
Nour MM, Dahoun T, Schwartenbeck P, Adams RA, Fitzgerald TH, Coello C, Wall MB, Dolan RJ, Howes OD (2018) Dopaminergic basis for signaling belief updates, but not surprise, and the link to paranoia. Proc Natl Acad Sci U S A 115:E10167-E10176.
O'Connell RG, Dockree PM, Bellgrove MA, Kelly SP, Hester R, Garavan H, Robertson IH, Foxe JJ (2007) The role of cingulate cortex in the detection of errors with and without awareness: a high-density electrical mapping study. Eur J Neurosci 25:2571–2579.
Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:156869.
O'Reilly JX, Schüffelgen U, Cuell SF, Behrens TE, Mars RB, Rushworth MF (2013) Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. Proc Natl Acad Sci U S A 110:E3660–E3669.
Orr C, Hester R (2012) Error-related anterior cingulate cortex activity and the prediction of conscious error awareness. Front Hum Neurosci 6:177.
Parr T, Friston KJ (2017) The active construction of the visual world. Neuropsychologia 104:92–101.
Payzan-Lenestour E, Dunne S, Bossaerts P, O'doherty JP (2013) The neural representation of unexpected uncertainty during value-based decision making. Neuron 79:191–201.
Pekkonen E, Rinne T, Näätänen R (1995) Variability and replicability of the mismatch negativity. Electroencephalogr Clin Neurophysiol 96:546–554.
Preuschoff K, 't Hart BM, Einhäuser W (2011) Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. Front Neurosci 5:115.
Raisig S, Welke T, Hagendorf H, van der Meer E (2010) I spy with my little eye: detection of temporal violations in event sequences and the pupillary response. Int J Psychophysiol 76:1–8.
Reimer J, McGinley MJ, Liu Y, Rodenkirch C, Wang Q, McCormick DA, Tolias AS (2016) Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. Nat Commun 7:13289.
Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. Science 274:1926–1928.
Saffran JR, Johnson EK, Aslin RN, Newport EL (1999) Statistical learning of tone sequences by human infants and adults. Cognition 70:27–52.
Shelley AM, Ward PB, Catts SV, Michie PT, Andrews S, McConaghy N (1991) Mismatch negativity: an index of a preattentive processing deficit in schizophrenia. Biol Psychiatry 30:1059–1062.
Smith C, Squire LR (2005) Declarative memory, awareness, and transitive inference. J Neurosci 25:10138–10146.
Stefanics G, Kremláček J, Czigler I (2014) Visual mismatch negativity: a predictive coding view. Front Hum Neurosci 8:666.
Turk-Browne NB, Jungé JA, Scholl BJ (2005) The automaticity of visual statistical learning. J Exp Psychol Gen 134:552–564.
Turk-Browne NB, Isola PJ, Scholl BJ, Treat TA (2008) Multidimensional visual statistical learning. J Exp Psychol Learn Mem Cogn 34:399–407.
Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK (2009) Neural evidence of statistical learning: efficient detection of visual regularities without awareness. J Cogn Neurosci 21:1934–1945.
Urai AE, Braun A, Donner TH (2017) Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. Nat Commun 8:14637.
van Veen V, Carter CS (2002) The anterior cingulate as a conflict monitor: FMRI and ERP studies. Physiol Behav 77:477–482.
Vogt BA, Hof PR, Friedman DP, Sikes RW, Vogt LJ (2008) Norepinephrinergic afferents and cytology of the macaque monkey midline, mediodorsal, and intralaminar thalamic nuclei. Brain Struct Funct 212:465–479.
Wacongne C, Changeux JP, Dehaene S (2012) A neuronal model of predictive coding accounting for the mismatch negativity. J Neurosci 32:3665–3678.
Weible AP, Weiss C, Disterhoft JF (2007) Connections of the caudal anterior cingulate cortex in rabbit: neural circuitry participating in the acquisition of trace eyeblink conditioning. Neuroscience 145:288–302.
Yu AJ, Dayan P (2005) Uncertainty, neuromodulation, and attention. Neuron 46:681–692.
Zekveld AA, Koelewijn T, Kramer SE (2018) The pupil dilation response to auditory stimuli: current state of knowledge. Trends Hear 22:2331216518777174.
Zénon A (2017) Time-domain analysis for extracting fast-paced pupil responses. Sci Rep 7:41484.
Zénon A, Solopchuk O, Pezzulo G (2019) An information-theoretic perspective on the costs of cognition. Neuropsychologia 123:5–18.