



# Neural Mechanisms Determining the Duration of Task-free, Self-paced Visual Perception

Shira Baror<sup>1,2</sup>, Thomas J Baumgarten<sup>1,3</sup>, and Biyu J. He<sup>1</sup> 

## Abstract

■ Humans spend hours each day spontaneously engaging with visual content, free from specific tasks and at their own pace. Currently, the brain mechanisms determining the duration of self-paced perceptual behavior remain largely unknown. Here, participants viewed naturalistic images under task-free settings and self-paced each image's viewing duration while undergoing EEG and pupillometry recordings. Across two independent data sets, we observed large inter- and intra-individual variability in viewing duration. However, beyond an image's presentation order and category, specific image content had no consistent effects on spontaneous viewing duration across participants. Overall, longer viewing durations were associated with sustained enhanced posterior positivity and anterior negativity in the ERPs. Individual-specific variations in the spontaneous

viewing duration were consistently correlated with evoked EEG activity amplitudes and pupil size changes. By contrast, presentation order was selectively correlated with baseline alpha power and baseline pupil size. Critically, spontaneous viewing duration was strongly predicted by the temporal stability in neural activity patterns starting as early as 350 msec after image onset, suggesting that early neural stability is a key predictor for sustained perceptual engagement. Interestingly, neither bottom-up nor top-down predictions about image category influenced spontaneous viewing duration. Overall, these results suggest that individual-specific factors can influence perceptual processing at a surprisingly early time point and influence the multifaceted ebb and flow of spontaneous human perceptual behavior in naturalistic settings. ■

## INTRODUCTION

What determines the duration of our spontaneous engagement with the perceptual environment? From taking a walk to the omnipresent social media engagement, everyday perceptual behavior is predominantly spontaneous: It is task-free and self-paced by the observer (Baror & He, 2021). However, in most laboratory-based experiments perceptual behavior is constrained by task demands, its timing and duration predetermined by experimental design. This approach is highly successful in providing insights into the neural basis of human perception but is inherently limited in its generalizability to natural perceptual experiences, which involve a variety of competing and coordinating mechanisms, resulting in high temporal variability both within and across individuals.

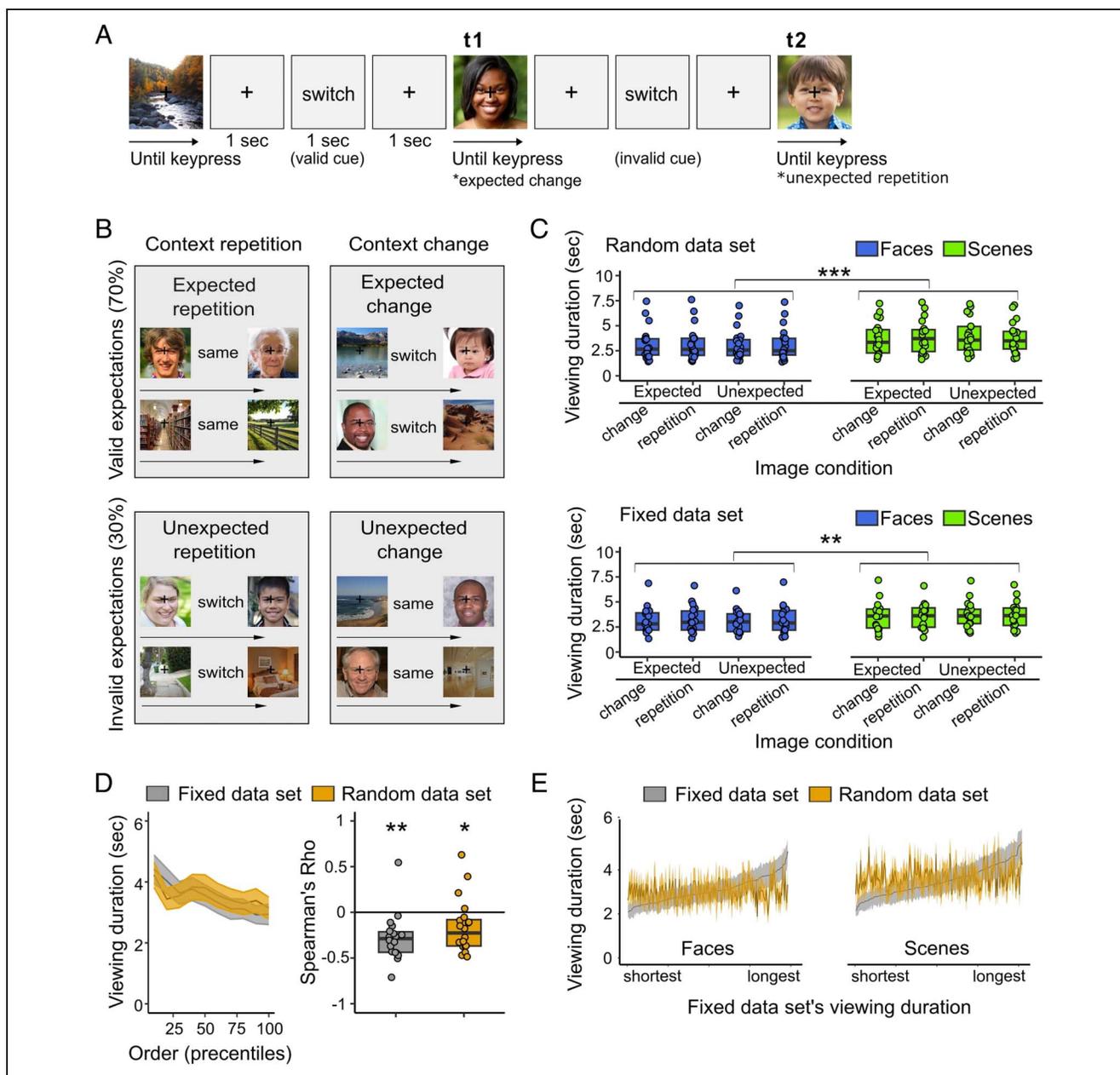
This fundamental gap is widely acknowledged in recent years, with the call for more ecologically valid studies that involve naturalistic stimuli (Snow & Culham, 2021; Haxby, Gobbini, & Nastase, 2020; Nastase, Goldstein, & Hasson, 2020; Sonkusare, Breakspear, & Guo, 2019) and encourage intersubject variability rather than suppress it (Miller et al., 2022). Accordingly, recent neuroscientific research on human perceptual behavior increasingly incorporates paradigms in which task constraints are minimized, such

as movie-viewing (Baldassano et al., 2017; Bartels & Zeki, 2004; Hasson, Nir, Levy, Fuhrmann, & Malach, 2004) or image free-viewing (Henderson, Goold, Choi, & Hayes, 2020; Henderson & Hayes, 2017). Nonetheless, in these earlier experiments, participants lacked agency over viewing durations and, consequently, inter- and intra-subject variability in the time domain is abolished. Thus, the underlying mechanisms of self-paced perceptual behaviors remain largely unexplored.

Here, we aimed to reveal the neural mechanisms determining the duration of task-free, self-paced visual perception. To this end, participants viewed a sequence of images, while only being asked to maintain eye fixation (to minimize EEG and eye-tracking artifacts) and press a key whenever they wanted to proceed to the next image (Figure 1A). In contrast to paradigms with fixed trial durations, this paradigm mimics the self-paced nature of naturalistic perceptual experiences such as social media engagement. Consequently, data contained significant intersubject as well as intrasubject variability in viewing durations. Participants' brain activity and pupil dynamics were measured by concurrent high-density EEG and high-speed eye-tracking during task performance.

Our secondary aim was to examine whether predictive processing influences the duration of self-paced perceptual behavior. Predictive processing has been shown to play a pivotal role in perception (de Lange, Heilbron, & Kok, 2018; Clark, 2013; Friston, 2005), facilitating

<sup>1</sup>New York University Grossman School of Medicine, <sup>2</sup>Hebrew University of Jerusalem, <sup>3</sup>Heinrich Heine University, Düsseldorf



**Figure 1.** Paradigm and behavior. (A) Images of faces and scenes were presented sequentially on the screen and viewing duration was self-determined by the participants. Cues were presented in between the images, with “switch” indicating a change of category (e.g., a scene is followed by a face) and “same” indicating category repetition (e.g., a face is followed by a face). Four hundred images were presented, divided into four blocks, allowing rest between blocks. (B) Image conditions. Cues were 70% valid, leading to four randomly ordered image conditions in each image category: expected repetition, expected change, unexpected repetition, and unexpected change. These conditions orthogonalize top-down expectations and bottom-up contextual repetitions. (C) Scene category effect on viewing duration. A three-way, repeated-measures ANOVA shows that in both data sets, scenes were viewed significantly longer than faces. Context and expectation conditions did not significantly influence viewing duration. (D) Serial order effect on viewing duration. Left: Mean viewing duration across participants as a function of the order of image presentation (40 trials in each bin). Shaded areas denote *SEM* across participants. Right: Serial order negatively correlated with viewing duration. Dots denote individual participants’ Spearman correlation values. (E) Mean viewing duration of images in the fixed data set ordered from shortest to longest according to group-mean (gray). Viewing duration of those images in the random data set (yellow) does not show a similar trend. Shaded areas depict *SEM* across participants. Boxes in the figure’s boxplot denote 50% of central data (between the first and third quartiles). Boxplot’s black lines indicate the median. \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

conscious perception, recognition, and categorization processes. However, self-paced perceptual behavior, such as viewing images on your phone, extends much longer durations than typically assessed in these earlier

paradigms. Whether the predictive influences that are typically found in temporally controlled settings with briefly presented stimuli extend to longer timescales in spontaneous viewing settings remains unknown.

To achieve our secondary aim, we included a trial-level cue (“same” or “switch”) that predicted an upcoming image category repetition or change with 70% validity (Figure 1B). Predictions regarding an upcoming event may be generated by learned bottom–up contextual regularities (Baumgarten et al., 2021) or top–down expectations (Kok, Mostert, & de Lange, 2017). Our paradigm orthogonalized these two factors, as contextual changes (or repetitions) could be expected or unexpected. This design allowed us to test whether top–down expectations and bottom–up context influence spontaneous viewing durations in a temporally extended, self-paced setting. On the basis of earlier findings, we hypothesized that both top–down and bottom–up predictive factors would influence spontaneous viewing duration.

To preview our main results, across two independent data sets, we found that the duration of spontaneous, self-paced visual perception for a specific image has large interindividual variability and is predicted by pupil size responses and evoked EEG magnitudes at an early latency (150 msec after image onset). Spontaneous viewing duration was also strongly predicted by the temporal stability in neural activity patterns starting at 350 msec from image onset. Interestingly, contrary to our hypothesis, neither bottom–up nor top–down predictive factors influenced spontaneous viewing duration, suggesting that perceptual behavior unfolding over longer timescales in a more naturalistic setting may be more robust to moment-to-moment predictive influences.

## METHODS

### Participants

This study involves two independent concurrent task-EEG/pupillometry data sets, with  $n = 20$  and  $n = 18$ , respectively. This sample size was determined by prior studies using similar methods, which typically have 25–35 participants in total. The two independent data sets provided a within-study replicability and generalizability check (see below for details on how the task protocol differed between them).

Thirty-eight participants took part in the experiment (19 women, mean age = 23.34 years), in two different data sets. The main analyses in the study were conducted using data from a group of 20 healthy participants, each viewing a randomly ordered sequence of face and scene images (hereafter referred to as the “random data set”); thus, all participants in this group viewed the same set of images but in different orders. Another group of 18 participants completed the experiment as a replication group. This data set differed from the main experimental group in that images were presented in a fixed order across participants (hereafter referred to as the “fixed data set”). Results of this replication group are reported alongside the main results. All participants signed informed consent before participating in the study; the study was approved by New York University Langone institutional review board (Protocol No.: s15–01323).

### Stimuli

Two hundred eighty scene images were taken from the BOLD5000 data set (Chang et al., 2019), depicting an equal proportion of indoor and outdoor scenes. Two hundred eighty face images were taken from the Flickr Faces HQ data set (Karras, Laine, & Aila, 2021) and were resized to  $375 \times 375$  pixels, to fit the scene images’ size. Face images depicted an equal proportion of male and female faces. Faces’ age ranged from infants to elderly adults, while paying special attention to inclusivity in race representation. Each participant viewed 200 scene images and 200 face images from these data sets. Participants viewed each image once throughout the experiment.

### Experimental Design and Procedure

In the main perception stage of the experiment, participants viewed images at their own pace, free from temporal or task constraints. This stage of the experiment comprised 400 images, divided into four blocks. Each trial began with a 1-sec presentation of a fixation cross, followed by a 1-sec presentation of a “same” or “switch” cue. The cue was followed by another 1-sec fixation cross presentation, after which the image appeared at screen center until the participant pressed a key, or until 10 sec have passed. The fixation cross remained superimposed on the images at screen center to minimize eye movements. Participants were instructed to maintain fixation whenever the fixation cross is visible. This was designed to minimize EEG artifacts and enable the spatio-temporal pattern similarity (STPS) analysis (see STPS Analysis section below).

Half of the images repeated the preceding image’s category, and half were of a category change. In 70% of the trials, the cue was valid, leading to a predicted category repetition or change, and in 30% of the trials, the cue was invalid, leading to an unpredicted category repetition or change. In total, the experiment involved four image conditions in each image category (face/scene): unexpected category change, unexpected category repetition, expected category change, and expected category repetition. This was designed to dissociate top–down expectations from bottom–up sensory change. Participants were explicitly informed that the cue was presented to help them predict the upcoming image’s category, but that they are free to view the image for as long as they wish, once the image is presented. At the end of each block, participants were given a break and were asked to resume the experiment at their own pace. The main perception stage was followed by a memory stage, which is succinctly analyzed in the Appendix.

### Trial Exclusion Criteria

Participants had 10 sec to make their decision to move on to the next image, after which the screen automatically

transitioned to the next trial. Trials that reached this 10-sec boundary without a participant's response were excluded from all analyses (mean = 6.05 trials per person, range = 0–29). In addition, the first trial in each block was excluded from all analyses, as, by design, these trials did not belong to a specific context/expectations condition, but rather set the condition for the subsequent trial. This resulted in excluding four additional trials from each participant's data.

### Statistical Analyses Scheme

To achieve our primary aim—to uncover the neural underpinnings of self-paced perceptual behavior, we computed two sets of correlational analyses. First, we computed correlations between pupillary and EEG-related measurements and the serial order of images, which was found to influence viewing durations (see Results section). Considering that serial order is not normally distributed, Spearman correlations were computed. In parallel, we computed Pearson correlations between pupillary and EEG-related measurements and spontaneous self-paced viewing duration, after controlling for serial order by using a linear regression approach. All correlations were first computed at the within-individual level and were followed by group-level analyses. To achieve our secondary aim—to test whether context and expectations influence self-paced behavior and its neural correlates, within-subjects, two-way, repeated-measures ANOVA analyses were employed, implementing Context and Expectations as the independent variables and the neural and pupillary measurements as the dependent variables. Below, we elaborate each statistical analysis, as well as the steps taken to control for multiple comparisons where relevant, separately for the behavioral, pupillary, and EEG-related analyses.

#### Behavioral Analyses

The influence of image condition on viewing duration was assessed using a repeated-measures, three-way ANOVA across participants, with Image Category, Context, and Expectations as independent variables and viewing duration as the dependent variable.

The influence of serial order on viewing duration was assessed by computing Spearman correlations between serial order and viewing duration individually for each participant. The effects were evaluated both at the single-subject level (as the number of participants showing a significant correlation) and at the group level (by a Wilcoxon signed-ranks test on correlations against 0). In addition, a linear mixed-effects model was conducted to assess the influence of serial order on viewing duration.

To test for consistency in content-specific effects between the fixed and random data sets, images were first sorted according to their group-mean viewing duration in

the fixed data set (divided to four groups: shortest, short, long, and longest). We then assigned each image in each of the four viewing-duration groups its mean viewing duration across participants in the random data set, in which image order was randomized across participants. This was followed by a one-way ANOVA examining whether these image groups are significantly different from one another in viewing duration.

Two luminance-related analyses were performed as well, to control for potential low-level effects. Using the SHINE (Spectrum, Histogram, and Intensity Normalization and Equalization) toolbox (Dal Ben, 2023), each image's luminance was computed. The correlation between luminance and mean viewing duration across participants for a particular image was computed. In addition, correlations in viewing duration across images between participants were computed, with the rationale that a consistent effect of luminance on viewing duration will result in high correlations between participants.

#### Eye-tracking Recording

Eye-tracking for both eyes was conducted using an SR Research Eyelink +1000 system (1000 Hz). To stabilize head position, a head post with chin and forehead rest was used. Nine-point calibration and validation were done at the beginning of each block as well as at the beginning of the memory stage.

#### Pupil Size Analysis

To characterize pupil size dynamics, all pupil size responses were first averaged across all participants. Evoked pupil size responses were baseline corrected to the 1000-msec prestimulus time window, in line with prior protocols (Oliva, 2019; Zhao et al., 2019; Zekveld, Heslenfeld, Johnsrude, Versfeld, & Kramer, 2014). Pupil size constriction started at ~300 msec after stimulus onset and was maximal between 700 and 900 msec, after which the pupil gradually dilated again. Considering these dynamics, subsequent analyses focused on baseline pupil size in the prestimulus time window (–1000:0 msec), the onset time window (700:900 msec), and the offset time window (–400:0 msec before offset), as well as on pupil size changes from prestimulus baseline, as measured at onset and offset. Because of trial-length variability and to prevent onset–offset overlap, we imposed an inclusion criterion such that only trials that were longer than 1500 msec were included in the analysis. Onset, offset, and prestimulus time-windows that had partially missing data because of blinks were excluded from the analyses. Correlations between pupil size measurements and spontaneous viewing duration or serial order were computed for each participant, followed by a Fischer Z transformation and a *t* test comparison against zero

at the group level. *p* Values were false discovery rate (FDR)-corrected for multiple comparisons.

### *EEG Recording*

One hundred twenty-eight Ag/AgCl actiCAP EEG electrodes (Brain Products GmbH) were placed according to the International 10–20 localization system. Four EOG electrodes were placed, two at the outer corner of each eye and two above and below the left eye. In addition, two reference electrodes were placed on the left and right mastoids. Skin was abraded using NuPrep skin gel before electrode placement. All electrodes were prepared with ABRALYT HiCl abrasive electrolyte gel (EASYCAP). Data were collected in DC recording mode using the BrainAmpDC system (Brain Products GmbH). Sampling rate was 1000 Hz. The experiment took place in a dark, electromagnetic interference-shielded, soundproof room.

### *EEG Preprocessing*

Data were preprocessed using the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). Raw data were first segmented at the block level, applying a 0.05-Hz high-pass filter, a 150-Hz low-pass filter, and a band-stop filter at 60 and 120 Hz to remove line noise. All filters were applied offline using a symmetrical third-order Butterworth filter with zero phase shift. This was followed by detrending, demeaning, and rereferencing the EEG data to the grand average. Electrodes were then manually inspected, and missing electrodes were interpolated using a neighboring approach. Basic detrending, demeaning, and rereferencing were done separately for EOG data after which EEG and EOG data were appended. Semi-automated jump and muscle artifact removal were carried out before running independent component analysis (ICA), which was applied to all EEG and EOG data. Subsequent artifact removal was done manually, inspecting for each participant the top number of components that explained > 90% of the variance. After removing ICA components considered to be artifacts, data were segmented into trials, from 1000 msec before stimulus onset to trial offset, and a low-pass filter at 35 Hz was applied to the epoched data, covering the delta-to-beta range. Faulty trials with discontinuous timeseries were inspected and removed. Subsequent analysis was done using custom written scripts in MATLAB (The MathWorks).

### *ERP Analysis*

To qualitatively compare ERP activity between short and long trials, ERPs were averaged across each participant's shortest 25% of trials and longest 25% of trials (using only trials longer than 1000 msec to avoid onset–offset overlap). Activity in each trial was baseline-corrected to the 500-msec prestimulus time window (Douglas, Maniscalco,

Hallett, Wassermann, & He, 2015). The time series of these upper and lower quartiles were averaged across participants, at trial onset (first 600 msec) and offset (last 400 msec). Whole-brain layout was divided into 12 spatial regions, in line with past work (Harel, Groen, Kravitz, Deouell, & Baker, 2016), and ERPs for each region are plotted.

Next, to characterize temporal dynamics of ERP's correlation with behavior, EEG activity was averaged across all trials in the onset and offset time windows, although incorporating only trials longer than 1000 msec to prevent onset–offset overlap. ERPs were then averaged in the onset and offset time windows in 50-msec time bins to obtain a less noisy measurement of neural activity. Correlations were computed separately for each participant, between ERP and absolute ERP magnitude at each time window/electrode and spontaneous viewing duration (Pearson correlation) or serial order (Spearman correlation). At the group level, correlation values at each electrode and timepoint were then Fisher *z*-transformed and compared against zero using a one-sample *t* test.

Subsequently, to examine the earliest latency of the correlation between spontaneous viewing duration and absolute ERP, ERPs were averaged across electrodes at each time bin and correlated with spontaneous viewing duration. At each time bin, correlations were Fisher *z*-transformed and compared against zero at the group level. Significance was evaluated by running a nonparametric cluster permutation test by shuffling the spontaneous viewing duration labeling at each iteration. *p* Value was calculated as the proportion of the randomized test statistic that exceeded the observed cluster's sum of statistic values (one-tailed test).

### *Time–frequency Analysis*

First, evoked power changes at image onset were characterized in the first 1000 msec of each trial and baseline-corrected using the  $-700:-200$ -msec prestimulus window. The 500-msec baseline window was shifted to end 200 msec before stimulus onset, to avoid temporal smearing between prestimulus and poststimulus activity (Min & Herrmann, 2007). This revealed a rapid increase in theta (4–8 Hz) power followed by a long-lasting alpha (8–13 Hz) power decrease in response to image onset, most prominent in posterior electrodes. On the basis of this onset response, we then tested whether the evoked power changes are modulated by spontaneous viewing duration or serial order, by correlating them with theta power (non-baseline-corrected) and theta power change (baseline-corrected; obtained from the 0:250-msec time window), alpha power and alpha power change (250:850-msec time window), and beta power and beta power change (15–25 Hz at the 250:850-msec time window). Each trial's power measurement was averaged across the relevant time window and frequency range across the 40 most posterior electrodes: P010, PP010h, PO8, TPP10h, P8, P6, TPP8h, TP10, CPP6h, P2, P4,

PP02h, PO4, P002, O2, OI2h, I2, POO10h, P09, PP09h, PO7, TPP9h, P7, P5, TPP7h, TP9, CPP5h, P1, P3, PP01h, PPO5h, PO3, P001, O1, OI1h, I1, POO9h, Pz, POz, and Oz.  $p$  Values were FDR-corrected for multiple comparisons.

### Power Spectrum Analysis

Power spectrum analysis was conducted in the 1000-msec prestimulus window of each trial. Correlations were computed between serial order or spontaneous viewing duration and power in each frequency bin and each electrode. At the group level, correlations were Fisher  $z$ -transformed, and a  $t$  test against zero was conducted.

### STPS Analysis

To estimate temporal neural stability at the trial level, single-trial, baseline-corrected ERPs were averaged in 50-msec time bins and correlations in whole-scalp neural activity pattern were computed between each time bin and all other bins in the trial, yielding an  $n$ -Bins by  $n$ -Bins STPS matrix. The size of each STPS matrix depended on the trial's length. Subsequently, an inclusion criterion was applied such that trials longer than 2500 msec were included in the following analyses.

We next used a 200-msec window sliding across the STPS matrices' diagonal from 0 to 2500 msec after image onset and averaged the correlations within the sliding window to obtain a "STPS index." For each participant, we calculated the correlation between the STPS index from each sliding window and spontaneous viewing duration. At the group level, correlations at each time window were Fisher  $z$ -transformed and were compared against zero. Significance was evaluated by a nonparametric cluster permutation test while shuffling the spontaneous viewing duration labeling at each iteration. Significance was calculated as the proportion of the randomized test statistic that exceeds the observed cluster's sum of statistic values. This revealed the cluster of timepoints in which the correlation between neural stability, as indexed in the STPS, and spontaneous viewing duration is significantly above chance.

In addition, for each participant, we computed the mean STPS for the onset (first 800 msec), middle (middle 800 msec), and offset (last 800 msec) parts of each trial. This matrix was averaged across trials for each participant. A group-level, repeated-measures, one-way ANOVA was run to examine differences in neural stability between the onset, offset, and middle trial parts. Subsequently, participants' across-trial mean STPS in each trial partition was correlated with their mean spontaneous viewing duration to examine if STPS could predict the overall propensity of a participant to view images longer or shorter.

## RESULTS

### The Duration of Spontaneous, Self-paced Perception Is Influenced by Serial Order and by High-level Image Category

Behavioral analysis revealed that participants in the random data set viewed scenes significantly longer than they viewed faces,  $F(19, 1) = 15.33$ ,  $\eta_p^2 = .44$ ,  $p < .001$  (Figure 1C, top), and this result was replicated in the fixed data set,  $F(17, 1) = 12.63$ ,  $\eta_p^2 = .42$ ,  $p < .002$  (Figure 1C, bottom). Beyond image category, whether the image changed in category from the previous image (i.e., bottom-up context change) and whether that category change or repetition was expected (i.e., top-down expectations) did not significantly influence viewing durations (context:  $p > .66$ ; expectations:  $p > .76$ ).

In addition to image category, viewing duration showed a significant negative correlation with serial order, such that images presented earlier in the experiment were viewed for longer durations. In the random data set, 17 out of 20 participants showed a significant correlation between serial order and viewing duration, 14 of which were negative (the correlation is also significant at the group level:  $p < .03$ , Wilcoxon signed-ranks test across participants; Figure 1D). In the fixed data set, 17 out of 18 participants showed a significant correlation between serial order and viewing duration, 16 of which were negative (group-level effect:  $p < .003$ , Wilcoxon signed-ranks test). Linear mixed-effects model of viewing duration by serial order likewise showed a significant effect (random data set:  $\beta = -1.65$ ,  $SE = 0.15$ ,  $t = -11.65$ ,  $p < 4.13e-31$ ; fixed data set:  $\beta = -3.38$ ,  $SE = 0.16$ ,  $t = -22.19$ ,  $p < 1.75e-105$ ).

We next examined whether specific image content, independently from image category, had a consistent effect on viewing durations across participants. To this end, we examined whether viewing durations of the images in the fixed data set were consistent with viewing durations of the same images in the random data set. Images in the fixed data set were sorted into four viewing-duration groups according to their group-mean viewing duration. We then tested whether viewing duration differed between these four groups in the random data set, which revealed a null result (faces:  $p > .99$  scenes:  $p > .97$ , one-way ANOVA; Figure 1E). This suggests that content-specific effects on viewing duration are idiosyncratic and differ between individuals.

A further control analysis examined whether luminance influenced viewing duration. We computed mean luminance of each image using the SHINE toolbox (Dal Ben, 2023). Overall, scenes' luminance ( $m = 45.23$ ,  $SE = 0.01$ ) was significantly lower than faces' luminance ( $m = 49.25$ ,  $SE = 0.1$ ;  $t = 5.96$ ,  $p = 4.2489e-09$ ). However, the correlation between image's viewing duration across participants and luminance was not significant for face images ( $\rho = -0.05$ ,  $p = .37$ ) or scene images ( $\rho = -0.08$ ,  $p = .15$ ). In addition, we computed correlations

in viewing duration across images between participants, for the reason that if luminance consistently influences viewing duration, viewing durations should be significantly correlated between participants across images. Mean rho values for correlations computed between each pair of participants was 0.02, both when computed across face images and when computed across scene images. Thus, luminance by itself does not appear to influence viewing duration.

### Spontaneous Viewing Duration Is Selectively Correlated with the Magnitude of ERPs

The behavioral analysis showed that beyond idiosyncratic observer-dependent factors likely related to the salience/significance of the specific image to the specific observer, self-paced viewing duration is influenced by serial order. In search for the neural mechanisms related to spontaneous self-paced behavior independent from serial order, we ran a regression analysis on viewing duration as a function of serial order and used the residuals of this regression as a proxy for spontaneous viewing duration with the influence of serial order removed. We first examined whether there is evidence showing that ERP activity is modulated by viewing duration. Next, we identified EEG activity features that correlated with either spontaneous viewing duration (hereafter referring to the regression residual—individual viewing duration unexplained by serial order) or with serial order (i.e., an image's sequence position within the overall experiment).

To qualitatively probe the potential relationship between viewing duration and ERP activity, we compared each participant's evoked responses in the shortest 25% and longest 25% of trials, and this comparison showed that a difference in ERP amplitudes between short and long trials emerged after ~300 msec from image onset. Although the very early ERP components (e.g., P1, N1, and P2) do not seem to be modulated by viewing duration, at subsequent timepoints, increased anterior negativity and increased posterior positivity are observed for long compared with short trials (Figure 2A). At trial offset, the difference in ERP amplitudes between short and long trials increased, primarily in anterior and posterior sites (Figure 2B). These qualitative observations demonstrate that longer viewing durations may be associated with stronger late ERP components at image onset, as well as with a slower return to baseline at image offset. Next, the relationship between viewing duration and ERP amplitudes was assessed on a trial-by-trial basis. For each participant, we computed the correlation between ERP amplitude, in each electrode and at each timepoint during onset and offset, and spontaneous viewing duration. This was followed by Fisher  $z$  transformation and a group-level  $t$  test comparison against zero, to test for significance at each timepoint and electrode (see Figure 2C for a schematic illustration). This showed that longer spontaneous viewing duration is associated with

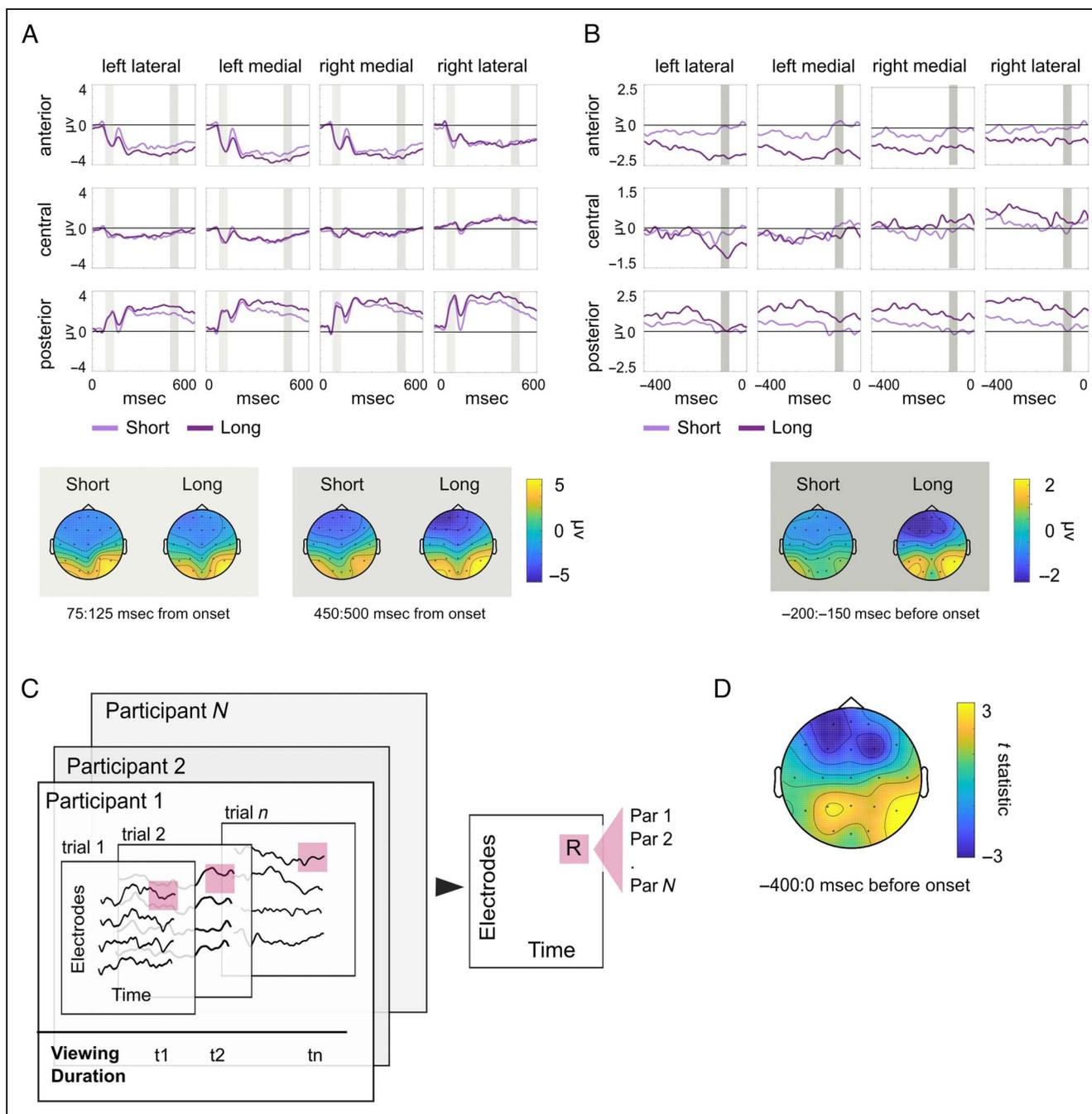
increased anterior negativity and increased posterior positivity at offset (the last 400 msec of each trial; Figure 2D).

As seen in Figure 2, an overall consistent posterior–anterior division was observed, with positive ERPs in posterior regions and negative ERPs in anterior regions, and the strength of the evoked responses changed over time. We next computed the absolute ERP strength, which succinctly summarizes the posterior–anterior activation pattern, and quantified its temporal dynamics (Figure 3A). The computation of absolute ERP strength is similar to computing “global field power” index (Zanesco, 2020), which reflects the electric field's power, yet is focused on amplitude rather than power.

For each participant, we computed the correlation between absolute ERP amplitude in each electrode and at each timepoint at the first 600 msec and at the last 400 msec of each trial, with serial order, and separately with spontaneous viewing duration. This analysis did not reveal a significant correlation between ERP amplitude and serial order (Figure 3B). However, absolute ERP amplitude significantly correlated with spontaneous viewing duration and this correlation gradually increased over time within a trial (Figure 3C). To reveal how early this correlation emerges at image onset, we averaged absolute ERP amplitude across all electrodes and examined its correlation with spontaneous viewing duration for each participant before conducting a population-level test (one-sample  $t$  test on Fisher  $z$ -transformed correlation values). This analysis revealed that a significant correlation between absolute ERP amplitude and spontaneous viewing duration emerged as early as 150 msec after image onset (random data set:  $p < .0001$ , cluster-based permutation test; Figure 3D; fixed data set:  $p < .0001$ ; cluster-based permutation test). Again, no correlation was found between serial order and absolute ERP amplitude. Lastly, this analysis was conducted again separately for face and scene images, to test for potential category-specific effects. The findings showed that the correlation between absolute ERP strength and spontaneous viewing duration significantly increases following image onset for both face and scene images, suggesting that this effect generalizes across image categories (Figure 3E).

### Serial Order Is Selectively Correlated with Alpha Power

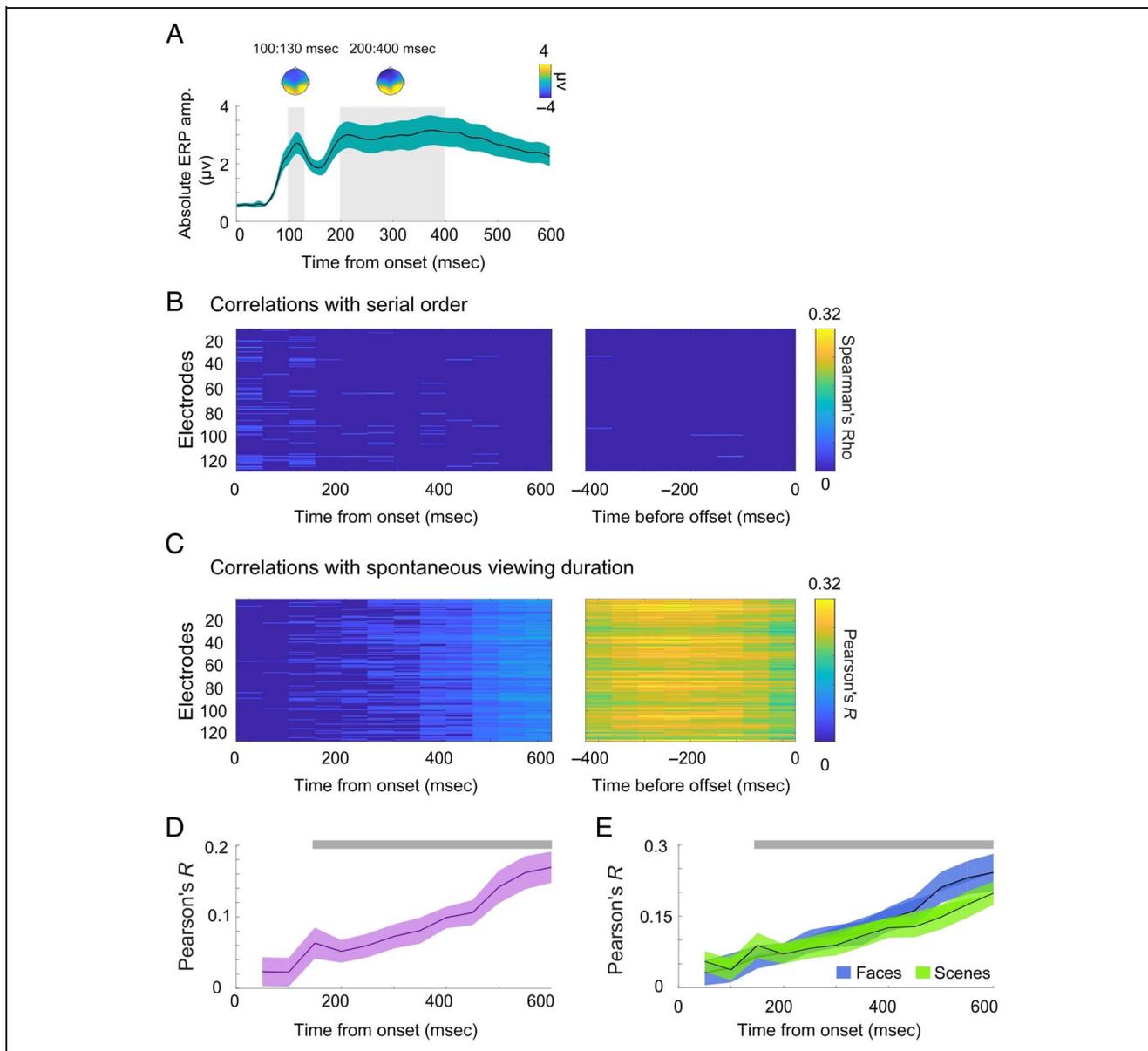
Next, we assessed the relationship between spontaneous viewing duration/serial order and EEG power changes. We first characterized band-limited power changes at trial onset, defined as the first 1000 msec after the onset of a new image. Grand-average time–frequency power changes (averaged across all trials, electrodes, and participants) revealed that image onset triggered an immediate increase in theta (4–8 Hz) power within 250 msec, followed by a long-lasting reduction in alpha power (8–13 Hz), and both features were most pronounced in posterior electrodes (Figure 4A).



**Figure 2.** ERP comparison between short and long trials: (A) Onset ERPs. Top: ERPs of short (shortest 25%, light purple) and long (longest 25%, dark purple) trials, averaged across participants over the first 600 msec of viewing duration, and divided to 12 spatial regions. Bottom: ERP topography of short and long trials in selected time windows (left: 75–125 msec from image onset, right: 450–500 msec from image onset). (B) Offset ERPs. ERPs of short (shortest 25%, light purple) and long (longest 25%, dark purple) trials, averaged across participants over the 400 msec preceding image offset, divided to 12 spatial regions. Bottom: ERP topography of short and long trials in a selected time window of 200–150 msec before image offset. The difference between short and long trials is primarily apparent before trial offset. \*ERPs are averaged across all trials longer than 1000 msec. Shaded gray areas indicate the selected time windows for topoplots. (C) Schematic illustration of trial-by-trial analysis of the correlation between ERPs and spontaneous viewing duration. At the participant level, ERPs at each time point and electrode (pink shaded region) were correlated with spontaneous viewing duration across trials. At the group level, at each time point and electrode, participants' array of correlations was Fisher  $z$ -transformed and compared against zero using a one-sample  $t$  test. (D) Topography of the  $t$  statistic of the group level  $t$  test comparison against zero, averaged across the offset's 400-msec time window.

Informed by this time–frequency domain characterization, we then examined whether spontaneous viewing duration and, separately, serial order modulated the onset-related power change response, by correlating

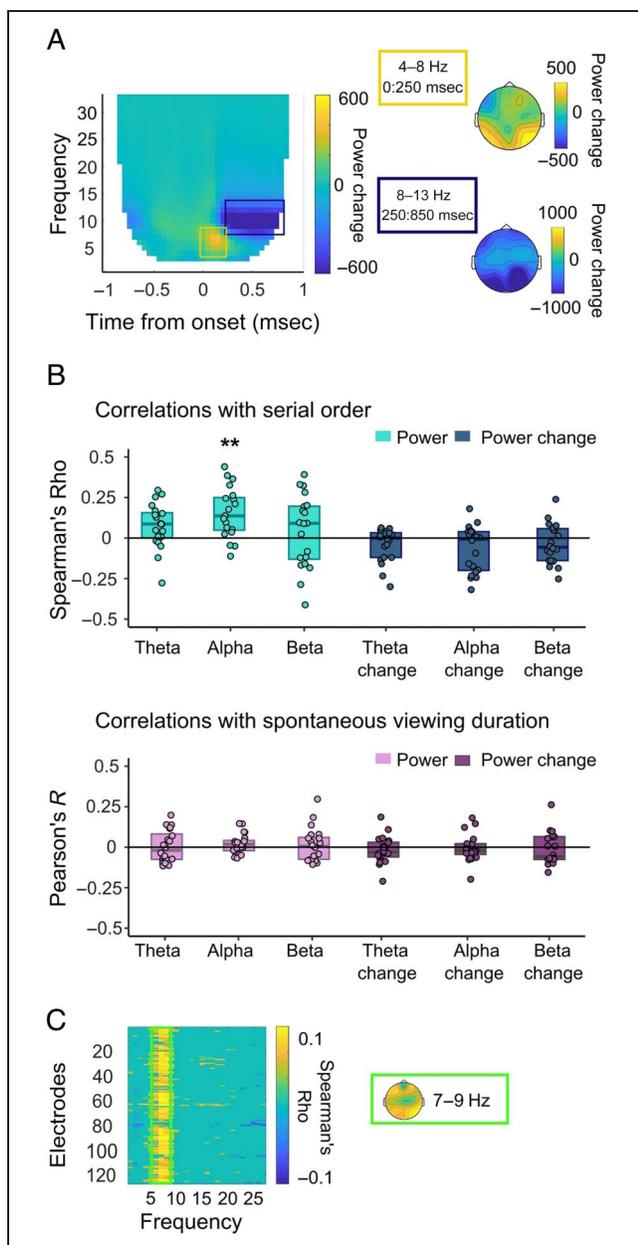
these two metrics with band-limited power in the theta (4–8 Hz, averaged in the time window 0–250 msec), alpha (8–13 Hz, 250–850 msec), and beta (15–25 Hz, 250–850 msec) ranges in posterior electrodes (see



**Figure 3.** Spontaneous viewing duration selectively correlates with absolute ERP amplitude. (A) Mean absolute ERP at onset averaged across all electrodes, trials, and participants. Shaded area depicts *SEM* across participants. Topographies depict the spatial distribution of ERP amplitude at each time window (indicated by gray shading), baseline corrected to the prestimulus 500-msec time window. ERPs are averaged across all trials longer than 1000 msec. (B) Group mean correlations between absolute ERP amplitude and serial order. (C) Group mean correlations between absolute ERP amplitude and spontaneous viewing duration. Correlations are computed over the onset (0:600 msec after image onset; left columns) and offset (−400:0 msec before image offset; right columns). (D) Individual correlations between absolute ERP at onset, averaged across all electrodes, and spontaneous viewing duration across all trials. (E) Individual correlations between absolute ERP at onset, averaged across all electrodes, and spontaneous viewing duration, separately for face (blue) and scene (green) images. In both D and E, shaded area depicts the *SEM* across participants. At the group level, correlations between absolute ERP and spontaneous viewing duration are significant as early as 150 msec after image onset (horizontal gray bar:  $p < .0001$ , assessed by cluster-based permutation tests).

Methods section for the list of posterior electrodes). This analysis showed that serial order significantly correlated with alpha power, such that greater alpha power was measured as the experiment progressed (random data set: alpha:  $\rho = 0.15$ ,  $t = 4.47$ ,  $p < .002$ , Figure 4B top; fixed data set: alpha:  $\rho = 0.23$ ,  $t = 6.61$ ,  $p < 4.41e-04$ ; FDR-corrected). Power in the theta and beta range correlated with serial order in the fixed data set (theta:  $\rho = 0.15$ ,

$t = 4.16$ ,  $p < .002$ ; beta:  $\rho = 0.20$ ,  $t = 3.36$ ,  $p < .008$ ; FDR-corrected) but not in the random data set (both  $p > .08$ ); as such, we do not emphasize this finding henceforth. The above results were obtained using raw power averaged within each frequency band and the respective time window. For image onset-related power changes (i.e., baseline corrected to a time window 200–700 msec before image onset), no significant



**Figure 4.** Serial order selectively correlates with baseline alpha power: (A) time–frequency analysis of the first 1000 msec from onset, baseline corrected to the  $-700$ : $-200$ -msec prestimulus time window, showing an immediate increase in theta power (4–8 Hz, yellow frame) followed by a long-lasting decrease in alpha power (8–13 Hz, blue frame). Right: topographies of onset-related power changes in the theta range, in the 0:250-msec window (yellow frame) and in the alpha range in the 250:850-msec time window (blue frame). (B) Correlations between power (i.e., non-baseline-corrected) as well as power changes (i.e., baseline-corrected to  $-700$ : $-200$ -msec prestimulus window) and serial order (top) or spontaneous viewing duration (bottom). On the basis of the finding in A, correlations were computed for theta (4–8 Hz at 0:250 msec), alpha (8–13 Hz at 250:850 msec), and beta (15–30 Hz at 250:850 msec) ranges. Boxes denote 50% of central data (between the first and third quartiles). Dark lines in the boxplots indicate the median.  $**p < .01$ , FDR-corrected. (C) Frequency-resolved correlation analysis between power in the 1-sec prestimulus time window and serial order, depicting significant correlations at the group level ( $p < .05$ , uncorrected). Topo plot (green frame) depicts rho values of the Spearman correlations between prestimulus activity in the 7- to 9-Hz range and serial order.

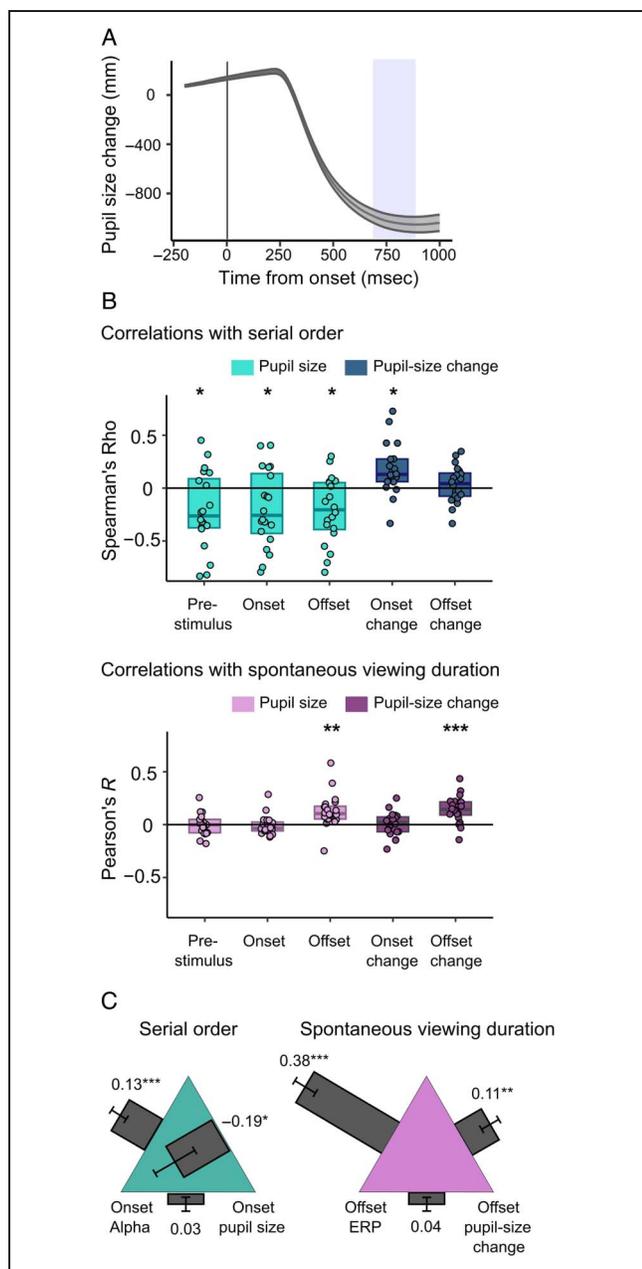
correlation with serial order was found. In contrast, spontaneous viewing duration did not correlate with any of the power metrics in either data set (results from the random data set shown in Figure 4B bottom).

We further probed whether the association found between serial order and alpha power is specific to image onset or rather a stimulus-independent process, by computing the power spectrum for each electrode in the 1-sec prestimulus time window (i.e., fixation period between cue offset and image onset; see Figure 1A), and correlating it with the image's serial order. This analysis revealed that prestimulus alpha power uniquely correlates with serial order before image onset, in a spatially nonspecific manner (Figure 4C). This result is consistent with prior observations showing a gradual increase in alpha power over the course of the experiment and extends these findings to naturalistic, task-free settings (Arnau, Brümmer, Liegel, & Wascher, 2021; Benwell et al., 2019).

Together, our EEG analyses point to a double dissociation between the neural correlates associated with idiosyncratic spontaneous viewing duration variations and those associated with serial order. Spontaneous viewing duration is correlated with ERP amplitudes as early as 150 msec after image onset, and serial order is selectively correlated with alpha power that gradually increases across the experiment. Although the alpha power finding is expected from previous literature, the finding that ERP amplitudes shortly after image onset predict spontaneous viewing duration (typically in the range of 2–5 sec; see Figure 1C) is surprising. The default explanation for such early neural effects is that they reflect bottom–up, image-related salience associated with low-level image features (such as contrast); however, if that were the case, spontaneous viewing duration should be correlated across participants, a scenario ruled out by our earlier behavioral and luminance analyses. Therefore, individual-specific, image-level effects can be seen in the evoked ERP responses, and the full mechanisms of this effect remain to be unveiled by future studies. One possibility is the influence of prestimulus spontaneous activity (McCormick, Nestvogel, & He, 2020; Podvalny, Flounders, King, Holroyd, & He, 2019; Baria, Maniscalco, & He, 2017; Sadaghiani, Poline, Kleinschmidt, & D'Esposito, 2015).

### Dissociable Pupillary Correlates of Spontaneous Viewing Duration and Serial Order

In parallel to the behavioral task and the EEG measurements, we tracked pupil size dynamics, with the aim to uncover the pupillary correlates of task-free, self-paced perceptual durations. We first characterized pupil response at image onset (Figure 5A). Pupil constriction started at  $\sim 300$  msec after stimulus onset, which was maximized at 700–900 msec (Figure 5A, blue shaded area), after which the pupil began to dilate again. Following this



**Figure 5.** Pupillary correlates of task-free, self-paced perception. (A) Pupil size response measured within the first 1 sec after image onset, baseline-corrected to the 1-sec prestimulus window. The black line depicts grand average, and the shaded area depicts SEM across participants. Pupil constriction is maximal at 700:900 msec after image onset (pale blue shading). (B) Pupil size metrics differentially correlate with serial order (top) and spontaneous viewing duration (bottom). Pupil size (bright boxplots) was measured in the prestimulus period (1000-msec time window before onset), at onset (between 700 and 900 msec after image onset) and at offset (from  $-400$  msec to 0 msec before image offset). Onset and offset pupil size changes (dark boxplots) were computed by baseline-correcting to the prestimulus time window. Boxes denote 50% of central data (between the first and third quartiles). Dark lines in the boxplots indicate the median. (C) Left: Partial correlations between serial order, alpha power at onset, and pupil size at onset. Right: Partial correlations between spontaneous viewing duration, absolute ERP at offset, and pupil size change at offset. Numbers indicate group-mean partial correlation values. Error bars depict SEM across participants. \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

characterization, we measured the correlations between serial order and, separately, spontaneous viewing duration with five pupillary metrics: baseline pupil size at the prestimulus time window (1 sec before image onset), pupil size at image onset (between 700 and 900 msec after onset), pupil size at image offset (from  $-400$  to 0 msec before stimulus offset), and pupil size change from the prestimulus time window as measured at image onset and image offset (see Methods section for details).

Serial order exhibited a robust negative correlation with pupil size measured at the prestimulus window, and during image onset and offset (random data set: prestimulus:  $\rho = -0.21, t = -2.58, p < .04$ ; onset:  $\rho = -0.19, t = -2.40, p < .04$ ; offset:  $\rho = -0.20, t = -2.93, p < .03$ ; Figure 5B top; fixed data set: prestimulus:  $\rho = -0.21, t = -2.56, p < .04$ ; onset:  $\rho = -0.18, t = -2.64, p < .04$ ; offset:  $\rho = -0.22, t = -3.46, p < .02$ ; FDR-corrected). However, image-elicited pupil size changes, measured either at onset or offset, did not exhibit a consistent relation with serial order (random data set: onset change:  $\rho = 0.17, t = 3.06, p < .03$ ; offset change:  $p > .39$ ; fixed data set: onset change:  $p > .12$ ; offset change:  $p > .79$ ; FDR-corrected).

By contrast, spontaneous viewing duration positively correlated with pupil size measured at offset whether corrected by prestimulus baseline or not (random data set: offset:  $r = .13, t = 3.48, p < .007$ ; offset change:  $r = .14, t = 4.92, p < .0006$ ; Figure 5B bottom; fixed data set: offset:  $r = .23, t = 3.48, p < 6.68e-04$ ; offset change:  $r = .28, t = 7.99, p < 3.65e-04$ , FDR-corrected). Spontaneous viewing duration did not correlate with any other pupil metrics with the exception of pupil size change measured at onset which, however, only showed a significant correlation in one of two data sets (fixed data set:  $r = .08, t = 3.42, p < .006$ ; random data set:  $p > .97$ ).

Together, these analyses point to another double dissociation between the mechanisms associated with serial order and those associated with spontaneous viewing duration: Although serial order is predominantly correlated with baseline pupil size irrespective of timing (whether before or after image onset), spontaneous viewing duration (independent of order) is primarily correlated with pupil size change from prestimulus baseline, measured at stimulus offset. The fact that the latter effect was evident at image offset but not onset is likely because of evoked pupil responses being relatively slow and taking time to develop.

### Neural Activity and Pupil Dynamics Independently Contribute to Serial Order and to Spontaneous Viewing Duration

Thus far, our findings show that spontaneous viewing duration (with the effect of serial order removed) is associated with ERP amplitudes and pupil size changes. We also found that serial order is associated with baseline alpha power and pupil size. We next sought to elucidate

whether the neural and pupillary correlates of each behavioral metric are distinct or associated with one another. To that aim, we conducted two partial-correlation analyses as described below. For each analysis, we chose EEG and pupil measures obtained in similar time windows and examined their correlation with behavior using partial correlations.

Partial correlation results concerning serial order are shown in Figure 5C, left. Serial order significantly correlated with alpha power measured at onset when controlling for pupil size measured at onset ( $r = .13, t = 4.37, p < 3.28e-04$ ). Similarly, serial order significantly correlated with pupil size when controlling for alpha power ( $r = -.19, t = -2.31, p < .04$ ). When controlling for serial order, onset pupil size and onset alpha power did not correlate with each other ( $p > .17$ ). These results were replicated in the fixed data set.

Partial correlation results concerning spontaneous viewing duration are shown in Figure 5C, right. Spontaneous viewing duration significantly correlated with absolute ERP amplitude measured at offset when controlling for pupil size change measured at offset ( $r = .38, t = 7.59, p < 3.58e-07$ ). Conversely, spontaneous viewing duration significantly correlated with pupil size change while controlling for ERP amplitude ( $r = .11, t = 3.00, p < .008$ ). When controlling for spontaneous viewing duration, offset ERP amplitude and offset pupil size change did not correlate with each other ( $p > .1$ ). These results were also replicated in the fixed data set.

Together, these analyses suggest that both for serial order and for spontaneous viewing duration, the neural correlates and pupillary correlates are independent from one another and thus likely represent different neural machinery involved. One possibility is that the pupillary correlates reflect the activity of subcortical neuromodulatory systems that our scalp-EEG recordings are less sensitive to.

### Neural Stability Predicts Spontaneous Viewing Duration

Inspired by previous work showing that longer viewing durations are associated with a static viewing style as measured in eye movement patterns (Zangrossi, Cona, Celli, Zorzi, & Corbetta, 2021), we predicted that longer viewing durations are associated with higher stability in neural activity patterns over time.

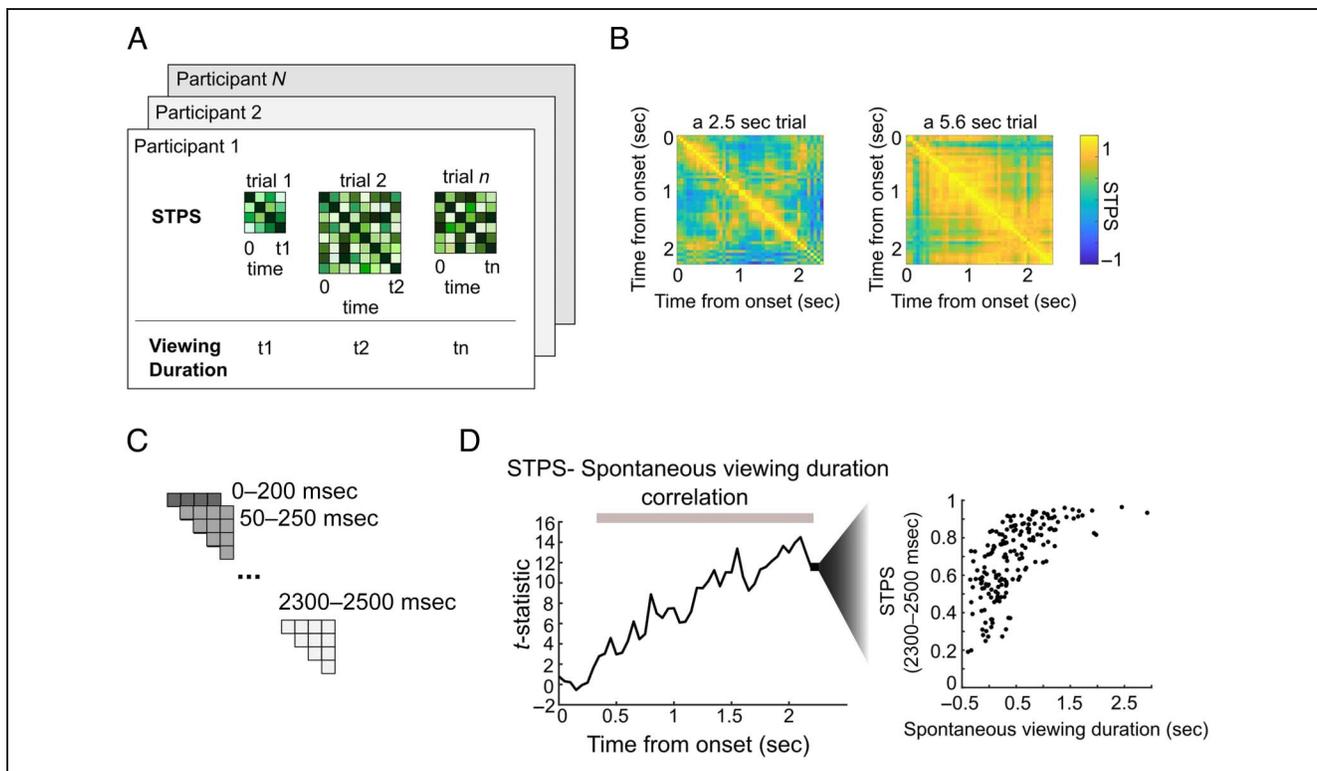
To test this prediction, we adapted the STPS analysis (Sols, DuBrow, Davachi, & Fuentesmilla, 2017). The STPS analysis typically focuses on finding the timepoints at which two events show the greatest neural similarity in their spatial activity patterns. This index has been used in recent years primarily in the realm of memory research showing that similarity in activity patterns across repeated presentations of a stimulus is associated with better memory of that stimulus and is referred to as quantifying representational stability across presentations (Sommer &

Sander, 2022; Baena, Cantero, & Atienza, 2021). Here, we computed the similarity in activity patterns between each pair of timepoints within a trial, with the rationale that greater similarity in neural activity patterns between consecutive timepoints indicates reduced neural change and greater neural stability.

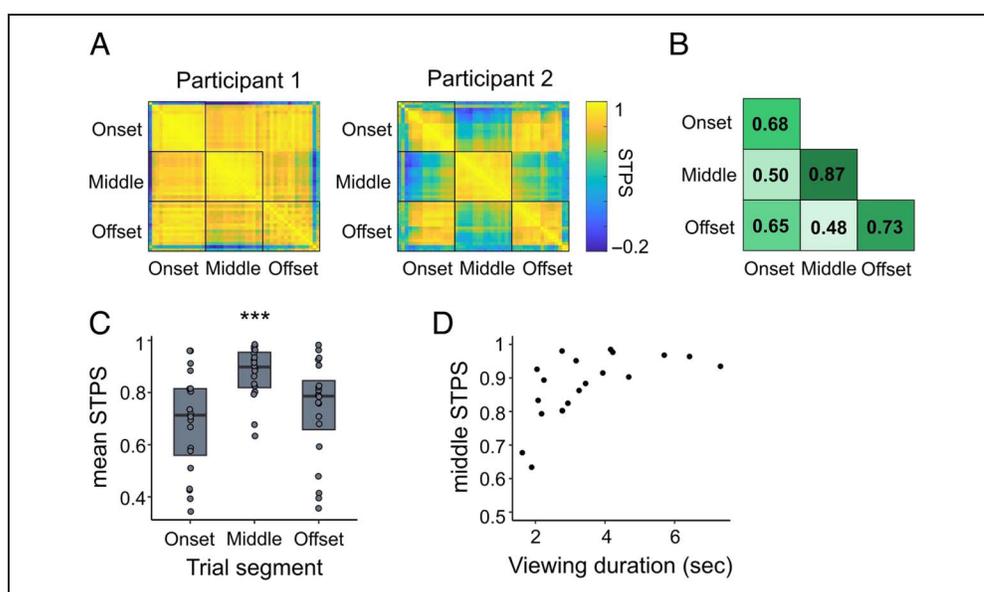
Briefly, we first computed the correlations between activity patterns at each time point and all other time points in a trial, generating a STPS matrix for each trial (Figure 6A and B). Then, a dynamic STPS index was computed using a 200-msec sliding window (Figure 6C). Subsequently, at every time window, we computed for each participant the correlation between spontaneous viewing duration and STPS index across trials. Given variable trial lengths, we used an arbitrary inclusion criterion by which only trials longer than 2500 msec were included in the analyses (corresponding to 58% of all trials; control analyses showed that this cutoff did not significantly impact the results). This allowed us to identify the specific time points at which STPS may be associated with spontaneous viewing duration. Example trials in Figure 6B demonstrate that over the same time window during image presentation, there is higher STPS shortly after image onset for longer compared with shorter trials.

This analysis revealed that STPS significantly correlated with spontaneous viewing duration as early as 350–550 msec after image onset ( $p < .0001$ , cluster-based permutation test; Figure 6D). In the fixed data set, this correlation emerged even earlier, at 200–400 msec after image onset ( $p < .0001$ , cluster-based permutation test). This correlation increased with time: Higher neural stability was increasingly indicative of longer viewing durations at later time points in the trial. This result suggests that after a brief visually evoked response within the first ~300 msec of image onset, the temporal stability of neural activity patterns is already a reliable index of the trial's eventual duration, with longer trials exhibiting greater stability compared with shorter trials.

We then sought to examine whether indices of neural stability also explain interindividual variability in spontaneous viewing durations. To that aim, we computed for each participant their cross-trial mean STPS at different times of a trial: the onset (i.e., first 800 msec of each trial), middle (i.e., middle 800 msec of each trial), and offset (i.e., last 800 msec of each trial; Figure 7A and B). A group-level, one-way ANOVA showed that middle-trial STPS was significantly higher than STPS measured in the onset and offset trial segments,  $F(18, 2) = 12.23, \eta_p^2 = .57, p < .001$  (Figure 7C); fixed data set:  $F(16, 2) = 16.36, \eta_p^2 = .49, p < .0001$ . In addition, middle-part STPS was significantly correlated with mean spontaneous viewing duration across participants, such that higher STPS was observed for participants who, on average, viewed the images for longer durations (Spearman's  $\rho = 0.63, p < .004$ ; Figure 7D). The latter result however did not reach significance in the fixed data set (Spearman's  $\rho = 0.27, p = .26$ ).



**Figure 7.** Between-participants differences in neural stability are associated with differences in spontaneous viewing duration: (A) examples of two participants' average STPS within and across non-overlapping onset (first 800 msec), middle (middle 800 msec), and offset (last 800 msec) segments of the trials. (B) Group mean STPS for each trial segment. (C) One-way ANOVA shows that middle-part STPS was significantly higher than STPS at onset or offset.  $***p < .001$ . (D) Scatter plot depicting the significant correlation between middle-part STPS and viewing duration across participants. Participants who exhibited longer viewing duration, on average, also exhibited greater middle-part STPS. This correlation was not found with onset or offset STPS.



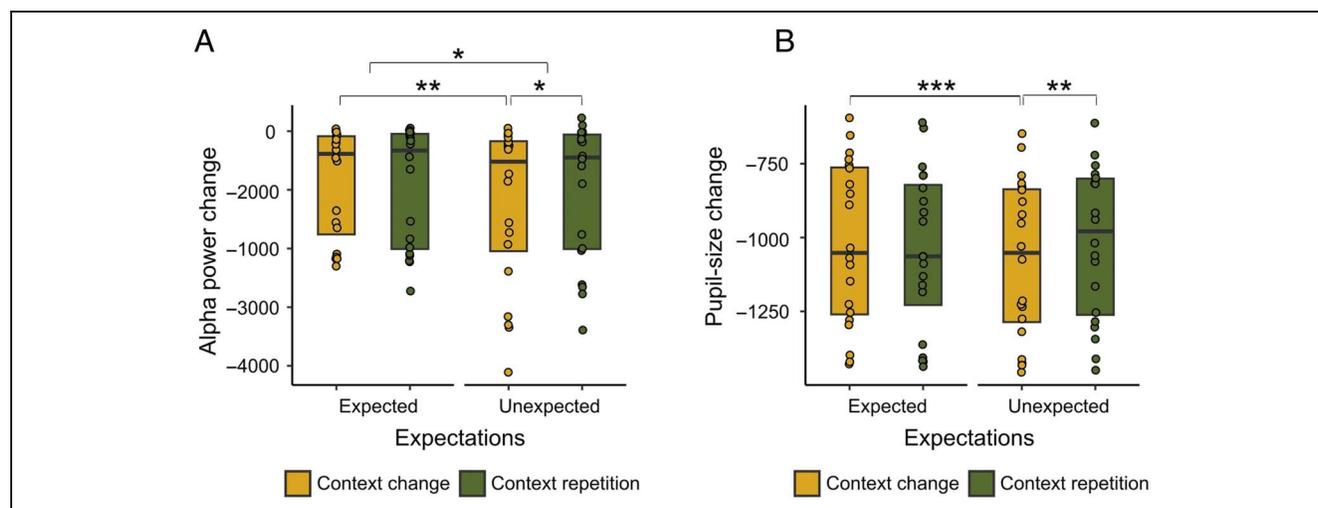
Together, these results show, for the first time to our knowledge, that neural stability predicts spontaneous self-paced perceptual behavior. Both within and across individuals, higher neural stability at an early latency (~350 msec after image onset) is associated with longer eventual viewing durations. Because spontaneous viewing durations investigated here are dissociated from serial order and have been shown to have idiosyncratic variations across individuals (see our earlier behavioral analysis), these results suggest that contextual factors unique to each individual (e.g., prestimulus spontaneous activity and memories) can influence neural activity at an early processing stage, to influence spontaneous perceptual behavior.

### Predictions Modulate Onset-related Neurophysiological Dynamics

Finally, to address our secondary aim, we tested whether bottom-up and top-down predictive factors influenced the duration of self-paced perceptual behavior and, if so, its corresponding neural correlates. As reported in the behavioral results section, bottom-up contextual changes as well as top-down expectations did not significantly influence spontaneous viewing duration (context:  $p > .66$ ; expectations:  $p > .76$ ; see Results section). Nevertheless, further analysis showed that expectations did modulate alpha power change at stimulus onset, such that a greater power decrease from baseline in the alpha range was measured for unexpected compared with expected stimuli,  $F(19, 1) = 6.13$ ,  $\eta_p^2 = .24$ ,  $p < .03$  (Figure 8A), consistent with earlier studies (Bridwell, Henderson,

Sorge, Plis, & Calhoun, 2018; Chao, Takaura, Wang, Fujii, & Dehaene, 2018; Rungratsameetaweemana, Itthipuripat, Salazar, & Serences, 2018; van Driel, Ridderinkhof, & Cohen, 2012). An additional interaction between context and expectation emerged,  $F(19, 1) = 6.53$ ,  $\eta_p^2 = .25$ ,  $p < .02$ , such that unexpected change, which involves lack of predictability from both bottom-up contextual change and top-down expectation violation, triggered the largest decrease in alpha power, unexpected change vs. unexpected repetition:  $t(19) = -1.96$ ,  $p < .04$ , Cohen's  $d = 0.44$ ; unexpected change vs. expected change:  $t(19) = -2.99$ ,  $p < .004$ , Cohen's  $d = 0.67$ . Analysis of onset-related pupil size change revealed similar results. Context and expectations interacted in their influence on pupil size change at onset,  $F(19, 1) = 10.67$ ,  $\eta_p^2 = .36$ ,  $p < .004$  (Figure 8B). The largest pupil size onset change (i.e., the greatest pupil constriction) was found under conditions of unexpected context change, unexpected change vs. unexpected repetition:  $t(19) = -2.82$ ,  $p < .005$ , Cohen's  $d = 0.63$ ; unexpected change vs. expected change:  $t(19) = -3.82$ ,  $p < .001$ , Cohen's  $d = 0.85$ .

Integrating our results here with past research showing that predictions take a central and early role in perception (Summerfield & de Lange, 2014; Kok, Brouwer, van Gerven, & de Lange, 2013; Bar, 2004), we propose that predictions may primarily be used for facilitating immediate image recognition and categorization processes (Hardstone et al., 2021; Press, Kok, & Yon, 2020; Kaiser, Quek, Cichy, & Peelen, 2019; O'Callaghan, Kveraga, Shine, Adams Jr., & Bar, 2017). At longer time-scales, however, a more complex and dynamic set of factors may come into play.



**Figure 8.** Predictions modulate onset-related neurophysiological changes. (A) Alpha power change measured at onset (mean across 8–13 Hz, at 250 msec:850 msec from image onset in posterior electrodes, baseline corrected to the  $-700$ : $-200$ -msec prestimulus time window) was significantly larger under conditions of unexpected (right bars) compared with expected (left bars) images. Unexpected change triggered the largest alpha-power decrease. (B) Pupil size change measured at onset (mean pupil size between at 700:900 msec from image onset, baseline corrected to the prestimulus 1000-msec time window) was significantly larger under conditions of unexpected change. Boxes denote 50% of central data (between the first and third quartiles). Black lines in the boxplots indicate the median. \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

## DISCUSSION

Everyday visual experiences are spontaneously self-paced by humans as agentic observers. Yet, most laboratory studies predetermine participants' percept durations, limiting the temporal variability that is inherent in naturalistic environments. Here, by allowing for individual temporal variability, we reveal several key findings regarding the neurophysiological mechanisms determining the duration of self-paced spontaneous image viewing. First, we found that viewing duration is heavily influenced by serial order and image category, but not significantly predicted by the specific image content at the group level or by image luminance. Second, stronger late ERPs, involving an anterior negativity and a posterior positivity, were found for long compared with short viewing durations. Importantly, after controlling for serial order, spontaneous viewing duration was strongly correlated with the amplitude of the ERP response and with pupil size change that develops slowly after image onset. These were doubly dissociable from the neurophysiological mechanisms associated with an image's serial order—baseline alpha power and baseline pupil size that slowly varied across the entire experiment. Third, as early as 350 msec after image onset, higher neural stability predicted longer spontaneous viewing duration. Finally, contrary to our prior hypothesis, predictive factors had no significant effects on spontaneous viewing durations in this task. Below, we discuss the main findings.

Serial order, the position of an image within the experiment, strongly influenced viewing durations. Serial order was associated with an increase in baseline alpha power, corroborating recent studies showing that compared with other frequency bands that remain stable over time, alpha power increases with time on task (Benwell et al., 2019). Serial order was also correlated with baseline pupil size, which decreased over time. The observed increases in alpha power and decreases in pupil size over the course of the experiment were nonspecific in time, manifesting in the prestimulus window, image onset, and image offset periods. Thus, we interpret them to signify slow changes in observer state, such as reduced novelty or decreased arousal with increased time on task.

Controlling for serial order allowed us to pinpoint the neural correlates underlying spontaneous variations in viewing duration. Spontaneous viewing duration dynamically correlates with the magnitude of the ERP response, such that stronger, late-latency ERPs were measured at onset for trials that would be viewed for longer. The correlation between ERP amplitude and spontaneous viewing duration increasingly strengthened toward image offset, where spontaneous viewing duration was correlated with increased anterior negativity and increased posterior positivity. Increased fronto-central negativity has been documented in the past, often attributed to the contingent negative variation (CNV; Walter, Cooper, Aldridge, McCallum, & Winter, 1964). The CNV is a negative centro-frontal component that emerges after several hundred

milliseconds from stimulus onset, and has been primarily associated with temporal expectancy (Boettcher, Stokes, Nobre, & van Ede, 2020; Walter et al., 1964) and preparatory attention (Funderud et al., 2012). However, the negative activity's topography observed here (see Figure 2D) is more anterior than the typically observed centro-frontal CNV topography (Damsma, Taatgen, de Jong, & van Rijn, 2020; Funderud et al., 2012; Ng, Tobin, & Penney, 2011). In addition, CNV topography typically has one anterior midline focus (Nobre & van Ede, 2018) rather than two bilateral foci observed here (Figure 2D). It is possible that a different, slowly evolving component belonging to the slow cortical potential (SCP) family (He & Raichle, 2009; Khader, Schicke, Röder, & Rösler, 2008; Birbaumer, Elbert, Canavan, & Rockstroh, 1990) contributes to the ERP findings depicted here, one that is associated more with agentic engagement, rather than with anticipating the duration of externally generated temporal intervals.

In parallel, considering that the peak amplitudes of early onset-related ERP components (e.g., P1) were not higher for longer compared with shorter viewing duration, extended agentic engagement with an image is likely sustained by a different neural process, reflected in the slower decay rate of the late-latency evoked neural activity.

Cognitively, the association between sustained ERP amplitudes and spontaneous viewing duration may also reflect changes in a deeper, more fine-grained or associative engagement with the image that takes place after the initial, swift gist-level processing (Baror & He, 2021; Campana, Rebollo, Urai, Wyart, & Tallon-Baudry, 2016). Importantly, because spontaneous viewing durations are not correlated across individual participants and are not correlated with image's luminance, these effects cannot be explained by purely bottom-up accounts.

Spontaneous viewing duration also correlated with the evoked pupil size response, and a partial correlation analysis revealed that this correlation was independent from the correlation with ERP amplitude. This suggests that pupil size changes signify different processes from those signified by ERP amplitudes in influencing spontaneous viewing duration. Transient pupil size changes have recently been implicated in evidence-accumulation processes, related to decisional components such as choice bias (de Gee et al., 2020; Urai, Braun, & Donner, 2017) or decision confidence (Colizoli, de Gee, Urai, & Donner, 2018). Therefore, these pupil size changes may reflect a decisional process of when to move onto the next visual content. It is important to note that for pupil size to correlate with behavior, there must be a mediating brain activity (e.g., subcortical arousal system, whose activity is more difficult to measure with EEG). It was previously shown that subcortical locus coeruleus/norepinephrine (LC/NE) neurons modulate pupil size changes (Joshi, Li, Kalwani, & Gold, 2016) and their phasic and tonic firing modes are associated with transient task-specific engagement and with overall arousal or exploration, respectively (Joshi

& Gold, 2020; Aston-Jones & Cohen, 2005). An intriguing possibility that remains to be tested is that serial order is related to tonic firing of LC/NE neurons, whereas spontaneous viewing duration is related to their phasic firing.

Interestingly, we found that spontaneous viewing duration was predicted by neural stability, indexed by increased similarity between neural activity patterns at consecutive timepoints. Neural stability correlated with spontaneous viewing duration, and this correlation increased with time, such that higher neural stability was increasingly more predictive of longer viewing duration at later time points in the trials. In a similar manner to the correlation found between ERP amplitude and spontaneous viewing duration, neural stability was predictive of viewing duration after a short delay, here as early as 350 msec after image onset. Given the latency of this correlation and the fact that it extends and increases over the time course of seconds (here measured up to 2.5 sec), the correlation between STPS and spontaneous viewing duration is likely to rely on SCPs (He & Raichle, 2009; Ergenoglu et al., 1998). The SCP is the recorded activity at the slow end of the field potential, mainly below 1 Hz, and although most ERP studies focused on early-latency components (< 600 msec), the SCP exhibits a later latency, starting from ~500 msec, and it lasts several seconds (Khader et al., 2008; Birbaumer et al., 1990). The SCP has been linked in the past with information integration across cortical areas, as well as with conscious perception and volition (He & Raichle, 2009), situating it as an optimal neural substrate to underlie the spontaneous behavior examined here, which extends several seconds, requires volitional self-paced behavior, and likely involves long-range information integration in the absence of a feature-limiting task.

Taking the ERP and STPS results together, it is suggested that task-free, self-paced perception is a dynamic process that begins with a brief evoked response that is agnostic to subsequent viewing duration (possibly related to initial gist-level processing) and that shortly thereafter, stronger cortical potentials as well as higher neural stability predict extended spontaneous viewing duration.

What factors predict the larger ERP amplitudes and higher neural stability (emerging at 350 msec), both of which associated with a longer eventual viewing duration (typically, 2–5 sec; 25th and 75th percentile: 1.8 and 4.8 sec)? Explanations for early-onset neural effects often resort to bottom-up factors. However, spontaneous viewing durations were not correlated across individuals (see behavioral results) and were not correlated with luminance, suggesting that bottom-up factors have limited impact on behavior. Rather, spontaneous viewing durations in response to the same visual content have large interindividual variability, and this variability likely comes from top-down factors such as related prior memories or image-unrelated spontaneous activations. A recent study showed that prior knowledge shapes the neural representations of images as early as 300 msec after image onset

(Flounders, González-García, Hardstone, & He, 2019). It is possible that such latent observer-specific prior memories also influence spontaneous viewing durations. Similarly, prestimulus spontaneous neural dynamics are known to influence image recognition processes (Podvalny et al., 2019; Sadaghiani et al., 2015) and may also influence spontaneous viewing duration.

The relation between increased stability and longer viewing duration aligns well with previous behavioral evidence showing that longer viewing durations are associated with a static rather than a dynamic viewing style, as characterized by saccadic patterns (Zangrossi et al., 2021). An exciting possibility is that the association between neural stability and image viewing duration may generalize to inform other naturalistic, self-paced behaviors, such as spontaneous social engagements, in which neural stability at an early latency may predict the duration of engagement. More generally, this finding opens the door to understanding the transitions between mental states through examining spontaneous neural dynamics. For example, recent findings show that spontaneous brain states can be identified by distinct neural patterns (Kucyi et al., 2021; Yamashita, Rothlein, Kucyi, Valera, & Esterman, 2021). An intriguing possibility is that an early stability index is predictive of the duration the brain “spontaneously” spends in a specific state and that changes in neural stability are indicative of an upcoming transition to a different state. If so, this would illuminate how mental events during spontaneous cognition (e.g., mind wandering) may be temporally structured without necessitating explicit retrospective report.

Although the current study makes a methodological advance in allowing participants to self-pace their viewing experience, a significant limitation is that it did not allow for free eye movements. This choice was made to facilitate neurophysiological and eye-tracking recordings, although we acknowledge that it interfered with spontaneous eye movements that would have emerged in fully naturalistic settings. Future studies that allow spontaneous eye movements during self-paced perception would be important for testing and extending our findings to fully naturalistic conditions.

Finally, our null behavioral findings about the impact of predictive factors on spontaneous viewing duration, despite positive findings of predictive factors’ influence on neurophysiological activity at image onset, beg additional investigation by future studies. One possibility is that these early neural effects (within 1 sec after image onset) are counteracted by later or more powerful processes that reflect an individual’s long-term preferences. Importantly, the predictive factors were manipulated at the level of image category, which was repeated or altered (bottom-up prediction) and expected or not (top-down prediction). Although image category itself had robust influences on viewing duration (Figure 1C), predictive factors did not. This result was contrary to our original hypothesis and highlights the importance for future studies to carefully outline the scope of predictive factors’ behavioral influences. It

seems plausible that predictive factors are more impactful on behavior in more difficult, threshold-level conditions (as often tested in prior studies).

In summary, perceptual experiences that are both task-free and self-paced are ever more prevalent, with the rise of social media and increased consumption of accessible online visual contents. In two independent data sets, we revealed replicable neurophysiological mechanisms that are involved in such naturalistic and agentic perception. Our findings reveal the mechanisms that selectively influences spontaneous viewing duration, which is highly idiosyncratic across individuals. These findings shed light on the complex brain mechanisms of spontaneous perceptual behavior and may inform future research that aims to uncover the neural basis of spontaneous behaviors and mental states in their naturalistically evolving contexts.

## APPENDIX

### Memory of Spontaneously Viewed Images Is Associated with Pupil Size Dynamics during Perception

At the end of the experiment, we conducted a surprise memory recognition test using a subset of the images that were presented during the perception stage. Participants were asked to decide as fast as possible whether each presented image appeared during the main perception stage

of the experiment or not. Memory performance was above chance across participants, with an average of 70% hit rate (std = 15%) and an average of 84% correct rejection rate (std = 10%). Mean sensitivity ( $d'$ ) was 1.65 (std = 0.56), and the mean measured criterion was 0.24 (std = 0.29). A paired  $t$  test comparing scene images' hit rate and face images' hit rate revealed that scenes were not remembered significantly better than faces ( $t = 1.82, p > .08$ ), despite being viewed longer at the perception stage.

Next, a mixed-effects logistic regression model was used to model memory accuracy, measured by hit rate. The model employed the main behavioral, neural, and pupillary parameters measured at the perception stage as fixed effects: serial order, spontaneous viewing duration (i.e., residuals), alpha power, offset absolute ERP, pupil size at onset, and pupil size change at offset. Participants were implemented as a random effect. The model is as follows:

$$P(\text{HIT}_{ij} = 1) \sim \sum_{z=0}^N \beta_z X_{zij} + \gamma_i + \epsilon_{ij} \quad (1)$$

Here,  $P$  denotes memory hit probability.  $z$  denotes the fixed effects parameters.  $\beta$  denotes the parameter's coefficient, and  $X$  denotes each parameter's measured input.  $i$  denotes individual participants, and  $j$  denotes trials.  $\gamma$  is the random effect parameter.

**Table 1.** Modeling Memory as a Function of Spontaneous Perception-related Dynamics

Model	$\beta$	SE	$t$	$p$ Value	CI
<i>Mixed-effects Logistic Regression Modeling of Memory Accuracy (Hit/Miss)</i>					
Serial order	0.0014	0.0008	1.6415	.101	[-0.0002, 0.003]
<b>Residuals</b>	<b>1.74E-04</b>	<b>7.52E-05</b>	<b>2.313</b>	<b>.02</b>	<b>[2.63E-05, 0.0003]</b>
Alpha power	7.53E-05	0.0001	0.6481	.517	[-0.0001, 0.0003]
Absolute ERP amplitude	-0.02001	0.0594	-0.6703	.5028	[-0.1565, 0.0768]
Onset pupil size	-0.0002	0.0002	-1.0737	.2832	[-0.0006, 0.0001]
Offset pupil size change	-0.0003	0.0001	-1.7752	.0762	[-0.0007, 3.70E-05]
<i>Mixed-effects Linear Regression Modeling of Memory RT</i>					
Serial order	-0.1107	0.1326	-0.8346	.4042	[-0.3713, 0.1498]
Residuals	-0.0105	1.05E-02	-1.0041	.3157	[-0.0312, 0.01]
Alpha power	2.87E-03	0.0153	0.1866	.852	[-0.0273, 0.033]
Absolute ERP amplitude	-3.9668	9.2454	-0.429	.668	[-22.129, 14.196]
Onset pupil size	0.0166	0.0297	0.5607	.5752	[-0.0417, 0.0751]
<b>Offset pupil size change</b>	<b>0.0627</b>	<b>0.0298</b>	<b>2.1004</b>	<b>.0361</b>	<b>[0.004, 0.122]</b>

Top: mixed-effects logistic regression modeling of memory accuracy (hit/miss). Bottom: mixed-effects linear regression modeling of memory RT in hit trials. Each model incorporated participants as a random factor and the six main parameters measured in the perception stage as the fixed factors. These factors include serial order, spontaneous viewing duration (i.e., residuals of the regression of viewing duration when controlling from order), alpha power at image onset, mean absolute ERP amplitude measured at offset, pupil size at image onset, and pupil size change at image offset. Significant predictors are shown in **bold**.

This analysis revealed that only spontaneous viewing duration significantly predicted memory accuracy (Table 1, top, shown in bold).

In addition, a mixed-effects linear regression model of memory RT in hit trials was conducted, using the same fixed-effect factors as above, and including participants as a random effect, as follows:

$$\text{memory RT}_{ij} \sim \sum_{z=0}^N \beta_z X_{zij} + \gamma_i + \varepsilon_{ij} \quad (2)$$

Here again,  $z$  are fixed-effects parameters,  $\beta$  denotes the parameter's coefficient, and  $X$  denotes each parameter's measured input.  $i$  denotes individual participant, and  $j$  denotes trials.  $\gamma$  is the random effect parameter.

This model revealed that pupil size change measured at image offset during the perception stage predicted the RT in the later memory task when the image is correctly identified as having been viewed before (Table 1, bottom).

### Acknowledgments

We thank Dillan Spector, Sebastian Sukdeo, and Jermaine Stokes for assistance with data collection, and Richard Hardstone and Jonathan Shor for discussions on the analysis.

Corresponding author: Biyu J. He, Neuroscience Institute, New York University Grossman School of Medicine, New York, NY 10016, or via e-mail: [biyu.he@nyulangone.org](mailto:biyu.he@nyulangone.org).

### Data Availability Statement

The analysis code and code to reproduce the figures can be downloaded at [https://github.com/BiyuHeLab/JOCN\\_Baror2024](https://github.com/BiyuHeLab/JOCN_Baror2024). Preprocessed, de-identified data are available from the corresponding author by reasonable request.

### Author Contributions

Shira Baror: Conceptualization; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing—Original draft; Writing—Review & editing. Thomas J. Baumgarten: Methodology. Biyu J. He: Conceptualization; Funding acquisition; Methodology; Project administration; Supervision; Validation; Writing—Original draft; Writing—Review & editing.

### Funding Information

This work was supported by a National Institutes of Health grant number: R01EY032085 and an Irma T. Hirschl Career Scientist Award (<https://dx.doi.org/10.13039/100006984>) to B. J. H.

### Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent

pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience (JoCN)* during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

### REFERENCES

- Arnau, S., Brümmer, T., Liegel, N., & Wascher, E. (2021). Inverse effects of time-on-task in task-related and task-unrelated theta activity. *Psychophysiology*, *58*, e13805. <https://doi.org/10.1111/psyp.13805>, PubMed: 33682172
- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450. <https://doi.org/10.1146/annurev.neuro.28.061604.135709>, PubMed: 16022602
- Baena, D., Cantero, J. L., & Atienza, M. (2021). Stability of neural encoding moderates the contribution of sleep and repeated testing to memory consolidation. *Neurobiology of Learning and Memory*, *185*, 107529. <https://doi.org/10.1016/j.nlm.2021.107529>, PubMed: 34597816
- Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, *95*, 709–721. <https://doi.org/10.1016/j.neuron.2017.06.041>, PubMed: 28772125
- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*, 617–629. <https://doi.org/10.1038/nrn1476>, PubMed: 15263892
- Baria, A. T., Maniscalco, B., & He, B. J. (2017). Initial-state-dependent, robust, transient neural dynamics encode conscious visual perception. *PLoS Computational Biology*, *13*, e1005806. <https://doi.org/10.1371/journal.pcbi.1005806>, PubMed: 29176808
- Baror, S., & He, B. J. (2021). Spontaneous perception: A framework for task-free, self-paced perception. *Neuroscience of Consciousness*, *2021*, niab016. <https://doi.org/10.1093/nc/niab016>, PubMed: 34377535
- Bartels, A., & Zeki, S. (2004). The chronoarchitecture of the human brain—natural viewing conditions reveal a time-based anatomy of the brain. *Neuroimage*, *22*, 419–433. <https://doi.org/10.1016/j.neuroimage.2004.01.007>, PubMed: 15110035
- Baumgarten, T. J., Maniscalco, B., Lee, J. L., Flounders, M. W., Abry, P., & He, B. J. (2021). Neural integration underlying naturalistic prediction flexibly adapts to varying sensory input rate. *Nature Communications*, *12*, 2643. <https://doi.org/10.1038/s41467-021-22632-z>, PubMed: 33976118
- Benwell, C. S. Y., London, R. E., Tagliabue, C. F., Veniero, D., Gross, J., Keitel, C., et al. (2019). Frequency and power of human alpha oscillations drift systematically with time-on-task. *Neuroimage*, *192*, 101–114. <https://doi.org/10.1016/j.neuroimage.2019.02.067>, PubMed: 30844505
- Birbaumer, N., Elbert, T., Canavan, A. G., & Rockstroh, B. (1990). Slow potentials of the cerebral cortex and behavior. *Physiological Reviews*, *70*, 1–41. <https://doi.org/10.1152/physrev.1990.70.1.1>, PubMed: 2404287

- Boettcher, S. E. P., Stokes, M. G., Nobre, A. C., & van Ede, F. (2020). One thing leads to another: Anticipating visual object identity based on associative-memory templates. *Journal of Neuroscience*, *40*, 4010–4020. <https://doi.org/10.1523/JNEUROSCI.2751-19.2020>, PubMed: 32284338
- Bridwell, D. A., Henderson, S., Sorge, M., Plis, S., & Calhoun, V. D. (2018). Relationships between alpha oscillations during speech preparation and the listener N400 ERP to the produced speech. *Scientific Reports*, *8*, 12838. <https://doi.org/10.1038/s41598-018-31038-9>, PubMed: 30150670
- Campana, F., Rebollo, I., Urai, A., Wyart, V., & Tallon-Baudry, C. (2016). Conscious vision proceeds from global to local content in goal-directed tasks and spontaneous vision. *Journal of Neuroscience*, *36*, 5200–5213. <https://doi.org/10.1523/JNEUROSCI.3619-15.2016>, PubMed: 27170119
- Chang, N., Pyles, J. A., Marcus, A., Gupta, A., Tarr, M. J., & Aminoff, E. M. (2019). BOLD5000, a public fMRI dataset while viewing 5000 visual images. *Scientific Data*, *6*, 49. <https://doi.org/10.1038/s41597-019-0052-3>, PubMed: 31061383
- Chao, Z. C., Takaura, K., Wang, L., Fujii, N., & Dehaene, S. (2018). Large-scale cortical networks for hierarchical prediction and prediction error in the primate brain. *Neuron*, *100*, 1252–1266. <https://doi.org/10.1016/j.neuron.2018.10.004>, PubMed: 30482692
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–204. <https://doi.org/10.1017/S0140525X12000477>, PubMed: 23663408
- Colizoli, O., de Gee, J. W., Urai, A. E., & Donner, T. H. (2018). Task-evoked pupil responses reflect internal belief states. *Scientific Reports*, *8*, 13702. <https://doi.org/10.1038/s41598-018-31985-3>, PubMed: 30209335
- Dal Ben, R. (2023). SHINE\_color: Controlling low-level properties of colorful images. *MethodsX*, *11*, 102377. <https://doi.org/10.1016/j.mex.2023.102377>, PubMed: 37771500
- Damsma, A., Taatgen, N., de Jong, R., & van Rijn, H. (2020). No evidence for an attentional bias towards implicit temporal regularities. *Attention, Perception, & Psychophysics*, *82*, 1136–1149. <https://doi.org/10.3758/s13414-019-01851-z>, PubMed: 31485992
- de Gee, J. W., Tsetsos, K., Schwabe, L., Urai, A. E., McCormick, D., McGinley, M. J., et al. (2020). Pupil-linked phasic arousal predicts a reduction of choice bias across species and decision domains. *eLife*, *9*, e54014. <https://doi.org/10.7554/eLife.54014>, PubMed: 32543372
- de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*, 764–779. <https://doi.org/10.1016/j.tics.2018.06.002>, PubMed: 30122170
- Douglas, Z. H., Maniscalco, B., Hallett, M., Wassermann, E. M., & He, B. J. (2015). Modulating conscious movement intention by noninvasive brain stimulation and the underlying neural mechanisms. *Journal of Neuroscience*, *35*, 7239–7255. <https://doi.org/10.1523/JNEUROSCI.4894-14.2015>, PubMed: 25948272
- Ergenoglu, T., Demiralp, T., Beydagi, H., Karamürsel, S., Devrim, M., & Ermutlu, N. (1998). Slow cortical potential shifts modulate P300 amplitude and topography in humans. *Neuroscience Letters*, *251*, 61–64. [https://doi.org/10.1016/S0304-3940\(98\)00498-4](https://doi.org/10.1016/S0304-3940(98)00498-4), PubMed: 9714465
- Flounders, M. W., González-García, C., Hardstone, R., & He, B. J. (2019). Neural dynamics of visual ambiguity resolution by perceptual prior. *eLife*, *8*, e41861. <https://doi.org/10.7554/eLife.41861>, PubMed: 30843519
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *360*, 815–836. <https://doi.org/10.1098/rstb.2005.1622>, PubMed: 15937014
- Funderud, I., Lindgren, M., Løvstad, M., Endestad, T., Voytek, B., Knight, R. T., et al. (2012). Differential go/nogo activity in both contingent negative variation and spectral power. *PLoS One*, *7*, e48504. <https://doi.org/10.1371/journal.pone.0048504>, PubMed: 23119040
- Hardstone, R., Zhu, M., Flinker, A., Melloni, L., Devore, S., Friedman, D., et al. (2021). Long-term priors influence visual perception through recruitment of long-range feedback. *Nature Communications*, *12*, 6288. <https://doi.org/10.1038/s41467-021-26544-w>, PubMed: 34725348
- Harel, A., Groen, I. I. A., Kravitz, D. J., Deouell, L. Y., & Baker, C. I. (2016). The temporal dynamics of scene processing: A multifaceted EEG investigation. *eNeuro*, *3*, ENEURO.0139-16.2016. <https://doi.org/10.1523/ENEURO.0139-16.2016>, PubMed: 27699208
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*, 1634–1640. <https://doi.org/10.1126/science.1089506>, PubMed: 15016991
- Haxby, J. V., Gobbini, M. I., & Nastase, S. A. (2020). Naturalistic stimuli reveal a dominant role for agentic action in visual representation. *Neuroimage*, *216*, 116561. <https://doi.org/10.1016/j.neuroimage.2020.116561>, PubMed: 32001371
- He, B. J., & Raichle, M. E. (2009). The fMRI signal, slow cortical potential and consciousness. *Trends in Cognitive Sciences*, *13*, 302–309. <https://doi.org/10.1016/j.tics.2009.04.004>, PubMed: 19535283
- Henderson, J. M., Goold, J. E., Choi, W., & Hayes, T. R. (2020). Neural correlates of fixated low- and high-level scene properties during active scene viewing. *Journal of Cognitive Neuroscience*, *32*, 2013–2023. [https://doi.org/10.1162/jocn\\_a\\_01599](https://doi.org/10.1162/jocn_a_01599), PubMed: 32573384
- Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour*, *1*, 743–747. <https://doi.org/10.1038/s41562-017-0208-0>, PubMed: 31024101
- Joshi, S., & Gold, J. I. (2020). Pupil size as a window on neural substrates of cognition. *Trends in Cognitive Sciences*, *24*, 466–480. <https://doi.org/10.1016/j.tics.2020.03.005>, PubMed: 32331857
- Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, *89*, 221–234. <https://doi.org/10.1016/j.neuron.2015.11.028>, PubMed: 26711118
- Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object vision in a structured world. *Trends in Cognitive Sciences*, *23*, 672–685. <https://doi.org/10.1016/j.tics.2019.04.013>, PubMed: 31147151
- Karras, T., Laine, S., & Aila, T. (2021). A style-based generator architecture for generative adversarial networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*, 4217–4228. <https://doi.org/10.1109/TPAMI.2020.2970919>, PubMed: 32012000
- Khader, P., Schicke, T., Röder, B., & Rösler, F. (2008). On the relationship between slow cortical potentials and BOLD signal changes in humans. *International Journal of Psychophysiology*, *67*, 252–261. <https://doi.org/10.1016/j.ijpsycho.2007.05.018>, PubMed: 17669531
- Kok, P., Brouwer, G. J., van Gerven, M. A. J., & de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *Journal of Neuroscience*, *33*, 16275–16284. <https://doi.org/10.1523/JNEUROSCI.0742-13.2013>, PubMed: 24107959
- Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences, U.S.A.*,

- 114, 10473–10478. <https://doi.org/10.1073/pnas.1705652114>, PubMed: 28900010
- Kucyi, A., Esterman, M., Capella, J., Green, A., Uchida, M., Biederman, J., et al. (2021). Prediction of stimulus-independent and task-unrelated thought from functional brain networks. *Nature Communications*, *12*, 1793. <https://doi.org/10.1038/s41467-021-22027-0>, PubMed: 33741956
- McCormick, D. A., Nestvogel, D. B., & He, B. J. (2020). Neuromodulation of brain state and behavior. *Annual Review of Neuroscience*, *43*, 391–415. <https://doi.org/10.1146/annurev-neuro-100219-105424>, PubMed: 32250724
- Miller, C. T., Gire, D., Hoke, K., Huk, A. C., Kelley, D., Leopold, D. A., et al. (2022). Natural behavior is the language of the brain. *Current Biology*, *32*, R482–R493. <https://doi.org/10.1016/j.cub.2022.03.031>, PubMed: 35609550
- Min, B.-K., & Herrmann, C. S. (2007). Prestimulus EEG alpha activity reflects prestimulus top-down processing. *Neuroscience Letters*, *422*, 131–135. <https://doi.org/10.1016/j.neulet.2007.06.013>, PubMed: 17611028
- Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: Rethinking the primacy of experimental control in cognitive neuroscience. *Neuroimage*, *222*, 117254. <https://doi.org/10.1016/j.neuroimage.2020.117254>, PubMed: 32800992
- Ng, K. K., Tobin, S., & Penney, T. B. (2011). Temporal accumulation and decision processes in the duration bisection task revealed by contingent negative variation. *Frontiers in Integrative Neuroscience*, *5*, 77. <https://doi.org/10.3389/fnint.2011.00077>
- Nobre, A. C., & van Ede, F. (2018). Anticipated moments: Temporal structure in attention. *Nature Reviews Neuroscience*, *19*, 34–48. <https://doi.org/10.1038/nrn.2017.141>, PubMed: 29213134
- O’Callaghan, C., Kveraga, K., Shine, J. M., Adams, R. B., Jr., & Bar, M. (2017). Predictions penetrate perception: Converging insights from brain, behaviour and disorder. *Consciousness and Cognition*, *47*, 63–74. <https://doi.org/10.1016/j.concog.2016.05.003>, PubMed: 27222169
- Oliva, M. (2019). Pupil size and search performance in low and high perceptual load. *Cognitive, Affective, & Behavioral Neuroscience*, *19*, 366–376. <https://doi.org/10.3758/s13415-018-00677-w>, PubMed: 30552642
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. <https://doi.org/10.1155/2011/156869>, PubMed: 21253357
- Podvalny, E., Flounders, M. W., King, L. E., Holroyd, T., & He, B. J. (2019). A dual role of prestimulus spontaneous neural activity in visual object recognition. *Nature Communications*, *10*, 3910. <https://doi.org/10.1038/s41467-019-11877-4>, PubMed: 31477706
- Press, C., Kok, P., & Yon, D. (2020). The perceptual prediction paradox. *Trends in Cognitive Sciences*, *24*, 13–24. <https://doi.org/10.1016/j.tics.2019.11.003>, PubMed: 31787500
- Rungratsameetaweemana, N., Itthipuripat, S., Salazar, A., & Serences, J. T. (2018). Expectations do not alter early sensory processing during perceptual decision-making. *Journal of Neuroscience*, *38*, 5632–5648. <https://doi.org/10.1523/JNEUROSCI.3638-17.2018>, PubMed: 29773755
- Sadaghiani, S., Poline, J.-B., Kleinschmidt, A., & D’Esposito, M. (2015). Ongoing dynamics in large-scale functional connectivity predict perception. *Proceedings of the National Academy of Sciences, U.S.A.*, *112*, 8463–8468. <https://doi.org/10.1073/pnas.1420687112>, PubMed: 26106164
- Snow, J. C., & Culham, J. C. (2021). The treachery of images: How realism influences brain and behavior. *Trends in Cognitive Sciences*, *25*, 506–519. <https://doi.org/10.1016/j.tics.2021.02.008>, PubMed: 33775583
- Sols, I., DuBrow, S., Davachi, L., & Fuentemilla, L. (2017). Event boundaries trigger rapid memory reinstatement of the prior events to promote their representation in long-term memory. *Current Biology*, *27*, 3499–3504. <https://doi.org/10.1016/j.cub.2017.09.057>, PubMed: 29129536
- Sommer, V. R., & Sander, M. C. (2022). Contributions of representational distinctiveness and stability to memory performance and age differences. *Aging, Neuropsychology, and Cognition*, *29*, 443–462. <https://doi.org/10.1080/13825585.2021.2019184>, PubMed: 34939904
- Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic stimuli in neuroscience: Critically acclaimed. *Trends in Cognitive Sciences*, *23*, 699–714. <https://doi.org/10.1016/j.tics.2019.05.004>, PubMed: 31257145
- Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*, 745–756. <https://doi.org/10.1038/nrn3838>, PubMed: 25315388
- Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, *8*, 14637. <https://doi.org/10.1038/ncomms14637>, PubMed: 28256514
- van Driel, J., Ridderinkhof, K. R., & Cohen, M. X. (2012). Not all errors are alike: Theta and alpha EEG dynamics relate to differences in error-processing dynamics. *Journal of Neuroscience*, *32*, 16795–16806. <https://doi.org/10.1523/JNEUROSCI.0802-12.2012>, PubMed: 23175833
- Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain. *Nature*, *203*, 380–384. <https://doi.org/10.1038/203380a0>, PubMed: 14197376
- Yamashita, A., Rothlein, D., Kucyi, A., Valera, E. M., & Esterman, M. (2021). Brain state-based detection of attentional fluctuations and their modulation. *Neuroimage*, *236*, 118072. <https://doi.org/10.1016/j.neuroimage.2021.118072>, PubMed: 33882346
- ZanESCO, A. P. (2020). EEG electric field topography is stable during moments of high field strength. *Brain Topography*, *33*, 450–460. <https://doi.org/10.1007/s10548-020-00780-7>, PubMed: 32514654
- Zangrossi, A., Cona, G., Celli, M., Zorzi, M., & Corbetta, M. (2021). Visual exploration dynamics are low-dimensional and driven by intrinsic factors. *Communications Biology*, *4*, 1100. <https://doi.org/10.1038/s42003-021-02608-x>, PubMed: 34535744
- Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., & Kramer, S. E. (2014). The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load. *Neuroimage*, *101*, 76–86. <https://doi.org/10.1016/j.neuroimage.2014.06.069>, PubMed: 24999040
- Zhao, S., Chait, M., Dick, F., Dayan, P., Furukawa, S., & Liao, H.-I. (2019). Pupil-linked phasic arousal evoked by violation but not emergence of regularity within rapid sound sequences. *Nature Communications*, *10*, 4030. <https://doi.org/10.1038/s41467-019-12048-1>, PubMed: 31492881