Stimulus Familiarity and Expectation Jointly Modulate Neural Activity in the Visual Ventral Stream

Mariya E. Manahova[1], Pim Mostert[1], Peter Kok[2], Jan-Mathijs Schoffelen[1], and Floris P. de Lange[1]

Abstract

■ Prior knowledge about the visual world can change how a visual stimulus is processed. Two forms of prior knowledge are often distinguished: stimulus familiarity (i.e., whether a stimulus has been seen before) and stimulus expectation (i.e., whether a stimulus is expected to occur, based on the context). Neurophysiological studies in monkeys have shown suppression of spiking activity both for expected and for familiar items in object-selective inferotemporal cortex. It is an open question, however, if and how these types of knowledge interact in their modulatory effects on the sensory response. To address this issue and to examine whether previous findings generalize to noninvasively measured neural activity in humans, we separately manipulated stimulus familiarity and expectation while noninvasively recording human brain activity using magnetoencephalography. We observed independent suppression of neural activity by familiarity and expectation, specifically in the lateral occipital complex, the putative human homologue of monkey inferotemporal cortex. Familiarity also led to sharpened response dynamics, which was predominantly observed in early visual cortex. Together, these results show that distinct types of sensory knowledge jointly determine the amount of neural resources dedicated to object processing in the visual ventral stream. ■

INTRODUCTION

Our visual environment is complex and rapidly changing, making visual perception a challenging task. To quickly parse visual input and deliver a stable percept of the world, the visual system is thought to employ principles of operation that allow an efficient sensory representation of the most likely current state of the visual world (Summerfield & de Lange, 2014; Meyer & Olson, 2011; Friston, 2005; Lee & Mumford, 2003; Rao & Ballard, 1999). By consequence, the amount of experience the visual system has had with a particular object can influence how much resources are allotted to processing that object. For instance, viewing an object image repeatedly (and thus becoming familiar with it) results in reduced spiking activity in inferotemporal cortex (IT) in monkeys (Miller, Li, & Desimone, 1991) and reduced hemodynamic activity in the human homologue (Grill-Spector, Henson, & Martin, 2006), lateral occipital cortex (LOC), as measured with fMRI. These findings suggest that familiar items require fewer neural resources than unfamiliar items.

Structure in visual information can also affect visual processing. If images are regularly presented in a specific temporal sequence, it becomes possible to predict which image will be presented next. Studies find that expected object images elicit reduced spiking activity compared with unexpected items in monkey IT (Kaposvari, Kumar, & Vogels, 2016; Meyer & Olson, 2011), whereas evidence is more mixed in human studies (Davis & Hasson, 2018; Egner, Monti, & Summerfield, 2010; Turk-Browne, Scholl, Johnson, & Chun, 2010; Puri, Wojciulik, & Ranganath, 2009), possibly due to differences in task demands (St. John-Saaltink, Utzerath, Kok, Lau, & de Lange, 2015).

Recently, Meyer, Walker, Cho, and Olson (2014) observed that image familiarization does not only lead to an activity reduction but also results in sharpening of the dynamics of neuronal visual responses in monkey IT. Because the sensory response to familiar images was truncated, this put IT neurons in a state of readiness for ensuing images and thereby enhanced their ability to track rapidly changing displays. This was demonstrated by a higher dynamic range (i.e., peak-to-trough difference) for familiar images compared with novel ones. A similar temporal truncation has been seen for expected, compared with unexpected, images (Meyer & Olson, 2011). Temporal sharpening and increased dynamic range may complement neural activity suppression in representing the visual world in a maximally efficient manner.

Although the effects of familiarity and expectation on the sensory response in monkey IT are relatively well described separately, it is uncertain whether and how these modulatory factors interact. Familiarity and expectation have, to date, been examined in distinct experimental paradigms, but because the two tend to go together (when we see a familiar image repeatedly, we

[1]Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen, [2]Yale University

come to expect it), their effects are difficult to disentangle. Moreover, these processes have not been investigated extensively in humans using electrophysiological measures, and evidence from noninvasive recordings in humans is mixed. It is important to note that there is a nested relationship between familiarity and expectation, such that we must be familiar with an object before we can expect to see it. Because of this, expected and unexpected items are by necessity familiar, but crucially expected and unexpected items can be equally familiar.

In the current study, we set out to study how image familiarity and expectation modulate the sensory response to object stimuli in humans. We did this by comparing the neural response to familiar versus novel images and within the class of familiar items by comparing the neural response to expected versus unexpected items. We measured neural activity using magnetoencephalography (MEG) while participants viewed images that were familiar or novel, and expected or unexpected. To preview, we found a reduction of activity for familiar compared with novel images in LOC. Within the class of familiar items, there was a further reduction of activity for expected compared with unexpected images in LOC. Moreover, we found a larger dynamic range for familiar compared with novel images that was most prominent in early visual areas, suggesting that the signal was temporally sharper for familiar than for novel input. These results show how familiarity and expectation jointly modulate activity in object-selective visual cortex, possibly allowing for an efficient coding of visual input.

METHODS

Participants

Twenty-nine healthy human volunteers (15 women, 14 men, mean age = 24.17 years, SD = 3.80 years) with normal or corrected-to-normal vision, recruited from the university's participant pool, completed the experiment and received either monetary compensation or study credits. The sample size, which was defined a priori, ensured at least 80% power to detect within-subject experimental effects with an effect size of Cohen's $d > 0.60$. The study was approved by the local ethics committee (CMO Arnhem-Nijmegen, Radboud University Medical Center) under the general ethics approval ("Imaging Human Cognition," CMO 2014/288), and the experiment was conducted in compliance with these guidelines. Written informed consent was obtained from each individual.

Stimuli

Stimuli were chosen from the image set provided at cvcl. mit.edu/MM/uniqueObjects.html. A different object was represented in each image, and all objects were shown against a gray background. A total of 2054 images were presented for each participant. Familiar images were randomly selected for each pair of participants. Each pair of participants saw different familiar images, and the images with and without a set sequence were counterbalanced within a pair of participants. Specifically, if for Participant 1, Set A comprised the images with a set sequence and Set B comprised the images without a set sequence, the opposite was true for Participant 2: Set B comprised the images with a set sequence, and Set A comprised the images without a set sequence. In both the behavioral and MEG sessions, the images subtended 4° of visual angle.

Apparatus

MATLAB (The Mathworks, Inc.) and the Psychophysics Toolbox extensions (Brainard, 1997) were used to show the stimuli on a monitor with a resolution of 1920 × 1080 pixels and a refresh rate of 100 Hz. For the MEG session, a PROpixx projector (VPixx Technologies) was used to project the images on the screen, with a resolution of 1920 × 1080 and a refresh rate of 100 Hz.

Experimental Design

On each trial of this experiment, participants saw a stream of images, and their task was to detect a target image. First, participants completed a behavioral training session in which they were familiarized with two sets of six images, one set with predictable structure and one set with no predictable structure. More specifically, they observed the images with a set sequence always in the same order, whereas the images with no set sequence were shown in a randomly shuffled order on each trial (Figure 1B). Importantly, the order for the images with a set sequence was circular, that is, each of the six images could be presented first. Images with a set sequence comprised 50% of trials, later referred to as "Familiar Sequence," and images with no sequence comprised the other 50%, later referred to as "Familiar No Sequence."

Participants performed a target detection task by pressing the spacebar when they saw an image of a rubber duck. Images of duckies were presented on 10% of trials as one of the six images in the sequence. The duckies were of eight different colors, and there were two viewpoints per color for a total of 16 images of duckies. Multiple images of duckies were used to reduce the possibility that participants may attend selectively to a particular color (e.g., yellow) or shape. The target task was chosen such that participants were required to maintain their attention on the visual stream.

During the behavioral training session, participants completed 10 blocks of 80 trials each for a total of 800 trials. Each block lasted 4.9 min, leading to a total training session duration of approximately 1 hr. At the end of the behavioral training session, participants' knowledge of the order of images with a set sequence was assessed with a sequence identification task. Participants were shown one of the six images from the sequenced set, and they

Figure 1. (A) Trial structure. First, participants saw a bull's eye fixation point for a jittered period between 500 and 750 msec. Then, each of the six images was presented for 180 msec with no gap between images. Finally, a fixation dot was presented at the end of the trial. (B, C, D) Experimental design. (B) Participants were familiarized with two sets of images. One set of images were always presented in a specific circular sequence (left), leading to Familiar Sequence Present trials. Another set of images were presented in a random sequence (right), resulting in Familiar No Sequence trials. (C) For the set of images with a set order, images were sometimes presented in the expected order (left), whereas on other occasions images were presented in an unexpected order (right). (D) New images were presented during Novel trials, and each unique image was only shown once during the whole experiment.
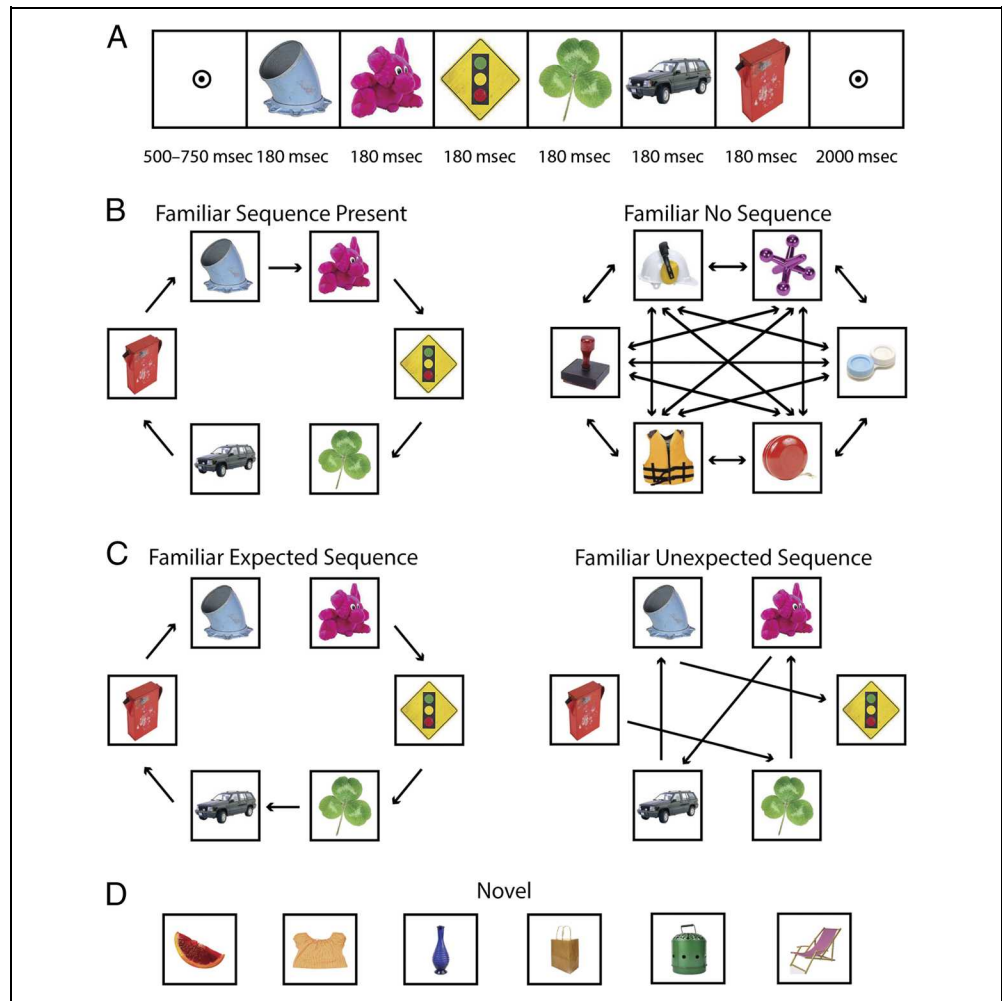
had to indicate which of the other five was most likely to follow it. This was done for each of the six images. The assessment took about 3 min.

One or two days later, participants completed the MEG testing session in which they saw familiar (i.e., those presented during the behavioral session) and novel (never seen before) images (see Figure 1B, C, and D for a depiction of the conditions). In contrast to the training session, the familiar images (sequence present) were now sometimes presented in the learned order (Expected, 50% of Familiar Sequence trials) and sometimes in a shuffled order (Unexpected, 50% of Familiar Sequence trials). The shuffled sequences for Unexpected trials were chosen in such a way that each image in the sequence was followed by an unpredicted image; in other words, none of the images were followed by the image they predicted (see Figure 1C). The Familiar No Sequence images were shown in shuffled orders, as during the training session. Familiar No Sequence images comprised one third of trials, and Familiar Sequence images also comprised one third of trials. The remaining third of trials was composed of novel images that the participants had not seen before (see Figure 1D). Unique novel im-

ages were used for every trial, so each novel image was only shown once during the experiment. To summarize, there were four conditions in the MEG session: Novel, Familiar No Sequence, Familiar Expected Sequence, and Familiar Unexpected Sequence (because the Familiar Sequence condition from the behavioral training session was divided into the Familiar Expected Sequence and Familiar Unexpected Sequence conditions).

Participants performed the same task as during the training session: They had to respond when they saw a ducky, and duckies were presented on 10% of trials. During the MEG testing session, participants completed eight blocks of 120 trials each for a total of 960 trials. There were 320 Familiar Sequence Present trials (160 Expected and 160 Unexpected), 320 Familiar No Sequence trials, and 320 Novel trials. Each block lasted 7.4 min, leading to a total experimental duration of approximately 1 hr. At the end of the MEG testing session, participants' knowledge of the familiar images was assessed. Participants saw 60 images, the 12 familiar ones and 48 selected at random from the novel images participants had been shown. Participants had to indicate whether the image was familiar or novel, where "familiar" referred to images seen repeatedly

during the behavioral training session as well as during the MEG testing session, whereas "novel" referred to images seen only once during the MEG testing session.

Trial Structure

In the behavioral training session as well as in the MEG testing session, each trial began with a fixation dot (see Figure 1A for the trial structure). The fixation dot was presented for a randomly selected period between 500 and 750 msec. Then, six images were shown, each lasting for 180 msec and presented back-to-back. At the end of a trial, if a target was presented during the trial and a response was given, the fixation dot turned green for 500 msec. If the response was incorrect, the fixation dot turned red for 500 msec. A response was considered incorrect on three occasions: if the participant pressed the button during a trial with a target stimulus but before the target was presented, if the participant pressed the button on a trial where no target was presented, or if the participant did not press the button on a trial where a target was presented. If no target was presented and no response was given, the fixation dot did not change color, and the white-and-black fixation dot remained on the screen for 750 msec. At the end of each trial, a blank screen was presented for 1250 msec, and participants were encouraged to blink during this period.

Data Acquisition

MEG Recordings

Brain activity was recorded using a 275-channel MEG system with axial gradiometers (VSM/CTF Systems) in a magnetically shielded room. During the experiment, head position was monitored online and corrected if necessary (Stolk, Todorovic, Schoffelen, & Oostenveld, 2013). This method uses three coils: one placed on the nasion, one in an earplug in the left ear, and one in an earplug in the right ear. To aid in the removal of eye- and heart-related artifacts, horizontal and vertical EOGs as well as an ECG were recorded. A reference electrode was placed on the left mastoid. The sampling rate for all signals was 1200 Hz. A projector outside the magnetically shielded room projected the visual stimuli onto a screen in front of the participant via mirrors. Participants gave their behavioral responses via an MEG-compatible button box. Participants' eye movements and blinks were also monitored using an eye-tracker system (EyeLink, SR Research Ltd.).

MRI Recordings

To allow for source reconstruction, anatomical MRI scans were acquired using a 3T MRI system (Siemens) and a T1-weighted MPRAGE sequence with a GRAPPA acceleration factor of 2 (repetition time = 2300 msec, echo time = 3.03 msec, voxel size = 1 × 1 × 1 mm, 192 transversal slices, 8° flip angle).

Data Analysis

Preprocessing of MEG Data

The MEG data were preprocessed offline using FieldTrip software (Oostenveld, Fries, Maris, & Schoffelen, 2011). Trials where target stimuli were presented and/or a response was given were removed from analysis. This was done because targets and responses elicited neural activity unrelated to the research question. Then, trials with high variance were manually inspected and removed if they contained excessive and irregular artifacts. This resulted in retaining, on average, 92% of trials per participant (range = 72–99%). Afterwards, independent component analysis was applied to identify regular artifacts such as heartbeat and eye blinks. The independent components for each participant were then correlated to the horizontal and vertical EOG signals and to the ECG signal. In this way, it was possible to identify which components most likely corresponded to the heartbeat and eye blinks. The data were filtered using a sixth-order Butterworth low-pass filter with a cutoff frequency of 30 Hz.

Event-related Fields

Before calculating event-related fields (ERFs), the data were baseline-corrected on the interval starting at 200 msec before stimulus onset until stimulus onset (0 msec). Subsequently, the data were transformed to simulate planar gradiometers to facilitate interpretation as well as averaging over participants. We applied a planar transformation and then a planar combination to the data. We took an equal number of trials per condition for each comparison to avoid any possible confounding influence of noise due to unequal number of trials. We did this by choosing a random selection of trials from the condition with more trials to match the number of trials in the condition with fewer trials. For the comparison between the Familiar No Sequence and Novel conditions, there were 320 trials per condition. For the comparison between the Familiar Expected Sequence and Familiar Unexpected Sequence conditions, there were 160 trials per condition.

Source Reconstruction on Time Domain Data

We performed source reconstruction to facilitate interpretation of the ERFs. Source reconstruction was done for 27 of the participants for whom we were able to acquire a structural MRI scan. We created volume conduction models based on a single-shell model of the inner surface of the skull and subject-specific dipole grids, which were based on a regularly spaced 6-mm grid in normalized MNI (Montreal Neurological Institute) space. For each grid point, the lead fields were rank reduced by removing the sensitivity to the direction perpendicular to the surface of the volume conduction model. Source activity was then obtained by estimating linearly constrained minimum variance spatial filters (van Veen, van Drongelen,

Yuchtman, & Suzuki, 1997), for which the data covariance was calculated over the interval of 200–1200 msec post-stimulus and regularized using shrinkage (as described in Blankertz, Lemm, Treder, Haufe, & Müller, 2011) with a regularization parameter of 0.01. The filters were applied to the axial gradiometer data and resulted in an estimated 2-D dipole moment per grid point, per time point.

For visualization as well as interpretation, we reduced these 2-D moments to a scalar value by taking the norm of the vector. This value reflects the degree to which a particular source location contributes to (differences in) activity measured at the sensor level. Critically, this value was obtained from the difference ERF between two conditions, rather than from each condition individually and subtracted afterward. In this way, differences in dipole orientation are also captured, instead of only magnitude, which would presumably correspond to different neural populations within the same source location.

One problem with taking the norm of the vector is that this is always a positive value and will therefore, due to noise, suffer from a positive bias. To counter this bias, we employed a permutation procedure, in which the condition labels were shuffled across trials. A total of 1000 permutations were performed, the average of which was taken as an estimate of the noise. Specifically, the average was calculated over the square of the dipole's norm (i.e., after squaring and summing in the Pythagorean theorem but before taking the square root). Next, this noise estimate was subtracted from the (square of) the true data, after which the data were divided by the noise estimate to counter the depth bias. The resulting values were then averaged over participants, and negative values were set to zero. Finally, the square root was taken, resulting in a group-level estimate of the contributions of each source location.

Spectral Analysis

To assess the dynamic range (i.e., the peak-to-trough difference) of the signal, we conducted a spectral analysis for all frequencies between 1 and 30 Hz. We applied the fast Fourier transform to the planar-transformed time domain data, after tapering with a Hanning window. The time period of interest was from 180 until 1080 msec, and data were baseline-corrected on the interval starting at 200 msec before stimulus onset until stimulus onset (0 msec). The spectral analysis was carried out separately per condition, and the resulting power per frequency was averaged over participants.

Source Reconstruction on Frequency Domain Data

We also applied source reconstruction analysis to facilitate interpretation of the power spectra. The source models and lead fields were obtained as described before and for the same 27 participants. Source activity was obtained by applying spatial filters based on partial canon-

ical correlations (Schoffelen, Oostenveld, & Fries, 2008) from the power data described above. The partial canonical correlation method allows for the efficient extraction of the source-level power for single trials. The regularization parameter was 0.01, and the frequency of interest was 5.6 Hz. This procedure resulted in an estimated 3-D dipole moment per grid point. For each grid point, we calculated the mean across each of the three spatial dimensions, computed its absolute value, and squared it. Then, we summed the resulting values for the three dimensions, which produced a single value per grid point. This analysis was carried out separately for each condition, and afterward we averaged the resulting values across participants.

Statistical Analysis

For the behavioral results, mean RT and accuracy were first calculated within participant per condition. Then, two-tailed paired-samples t tests were calculated for the two relevant conditions for a comparison. Behavioral data were analyzed for 28 of the 29 participants, as a technical issue in data acquisition prevented the analysis of behavioral data of the first participant.

To statistically assess the MEG activity difference between conditions in the time domain and control for multiple comparisons, we applied cluster-based permutation tests (Maris & Oostenveld, 2007), as implemented by FieldTrip (Oostenveld et al., 2011). The tests were carried out on the time period between 0 and 1200 msec, 0 msec being the onset of the first stimulus, over all sensors, and 10,000 permutations were used per contrast. For each sensor and time point, the MEG signal was compared univariately between two conditions, using a paired t test. Positive and negative clusters were then formed separately by grouping spatially and temporally adjacent data points whose corresponding p values were lower than .05 (two-tailed). Cluster-level statistics were calculated by summing the t values within a cluster, and a permutation distribution of this cluster-level test statistic was computed. The null hypothesis was rejected if the largest cluster in the considered data was found to be significant, which was the case if the cluster's p value was smaller than .05 as referenced to the permutation distribution. The SEM was computed using a correction that makes it suitable for within-subject comparisons (Morey, 2008; Cousineau, 2005).

We also applied cluster-based permutation tests to statistically assess MEG activity differences between conditions in the frequency domain (see Data Analysis). The tests were carried out on the $\log_{10}$-transformed data for the frequency of interest (stimulus presentation frequency: 5.6 Hz), over all sensors, and with 10,000 permutations per contrast. Adjacent sensors with nominal p values lower than .05 (two-tailed) were grouped into clusters. The t values within a cluster were summed, yielding a cluster-level statistic. If the largest cluster's p value was smaller

than .05, the difference across the compared conditions was considered statistically significant.

After having established a difference in neural activity between expected and unexpected stimuli, we wished to determine whether our data offered evidence for surprise enhancement or expectation suppression (or both). To this end, we compared Familiar Unexpected Sequence trials with Familiar No Sequence trials, as well as Familiar Expected Sequence with Familiar No Sequence trials. To match the number of trials per condition, we randomly selected 160 trials from the Familiar No Sequence condition to match the number of trials (160) in the Familiar Expected Sequence and Unexpected conditions. For these comparisons, the time domain data were averaged per condition over a time period from 500 until 900 msec after the onset of the visual sequence because this was identified as the time period of significant difference between expected and unexpected stimulus streams (see Results). Next, two paired-samples t tests were carried out to statistically compare the amplitude of Familiar Unexpected Sequence and Familiar No Sequence trials, as well as the amplitude of Familiar Expected Sequence and Familiar No Sequence trials.

To determine whether the expectation effect changed over the course of the experiment, we divided the data into two: the first half and the second half of the experiment. Then, we averaged over Familiar Expected Sequence trials and over Familiar Unexpected Sequence trials for each half. We computed the ERFs and applied the planar transformation and combination for each of the four conditions (Expected first half, Expected second half, Unexpected first half, and Unexpected second half). After that, we computed the difference for each half: Unexpected − Expected first half and Unexpected − Expected second half. We compared the two differences (Unexpected − Expected) for the first half and for the second half by calculating a paired samples t statistic and conducting a cluster-based permutation test. The test was run on the time period from 0.5 until 0.9 sec, as this was the time period identified by a previous analysis (see Results).

Finally, to assess whether explicit knowledge of the sequence immediately after training influenced the activity modulation by expectation, we divided participants based on whether their performance was above or below chance level (20%) on the sequence identification task. This division resulted in two groups of 14 participants each. We averaged each participant's data per condition from 500 to 900 msec and over the sensors that contributed to the significant difference for Expected versus Unexpected sequences. We examined whether there were differences between the groups using an independent-samples t test. In addition, we conducted a Spearman's rank correlation to determine whether there was a relationship between participants' performance on the sequence learning task and their expectation effects, where the expectation effect was quantified in the same way as for the independent-samples t test.

RESULTS

Behavioral Results

The participants' task was to press a button whenever they saw a target stimulus, in this case an image of a ducky. Participants were at near ceiling level in their performance on the task (mean accuracy = 94.9%, SD = 2.9%). Participants' accuracy was not significantly affected by whether duckies appeared in Familiar No Sequence versus Novel trials ($t_{27} = -0.75, p = .46$) or in Familiar Expected Sequence versus Familiar Unexpected Sequence trials ($t_{27} = 0.40, p = .69$). Furthermore, participants' RTs to target trials were not significantly different between duckies embedded in Familiar No Sequence versus Novel trials ($t_{27} = -1.74, p = .09$) or Familiar Expected Sequence versus Familiar Unexpected Sequence trials ($t_{27} = -0.04, p = .97$).

At the end of the behavioral training session, participants' knowledge of the order of the predictable images was assessed. On average, when participants were shown an image and had to report which image followed it, they selected the correct image 25% of the time (SD = 19.7%), with chance level performance at 20%. This suggests that subjects were largely unaware of the sequence, in agreement with their verbal reports.

At the end of the MEG session, participants' knowledge of image familiarity was assessed. On average, when participants had to report whether an image was familiar or novel, they correctly identified the familiar images in 91.9% of trials (SD = 5.8%), showing that they were clearly aware of the image familiarity manipulation.

MEG Results

Familiar Items Lead to Reduced Activity in LOC

To investigate the difference in amplitude between familiar and novel items without any influence of the expectation manipulation, we compared the Familiar No Sequence and Novel conditions because participants did not learn a sequence for the images in the Familiar No Sequence condition. A significant difference ($p < 1e-6$) across conditions was observed, which was driven by the cluster of sensors shown in Figure 2A from approximately 200 msec until 1200 msec. The black asterisks in the figure denote sensors that contribute to this cluster for at least half of the time period from 200 to 1200 msec. The average time course for the sensors contributing to the cluster is plotted in Figure 2C; the black line at the bottom shows that at least one of the selected sensors at that time point contributes to the significant difference. Clearly, novel items lead to significantly more activity than familiar ones. Source reconstruction revealed that the difference between familiar and novel items stemmed primarily from sources along the visual stream. This included early visual areas as well as downstream visual areas such as right and left inferior occipital gyrus (Figure 2B; Table 1) in the vicinity of lateral occipital complex (LOC; Grill-Spector, Kushnir, Hendler, & Malach, 2000).
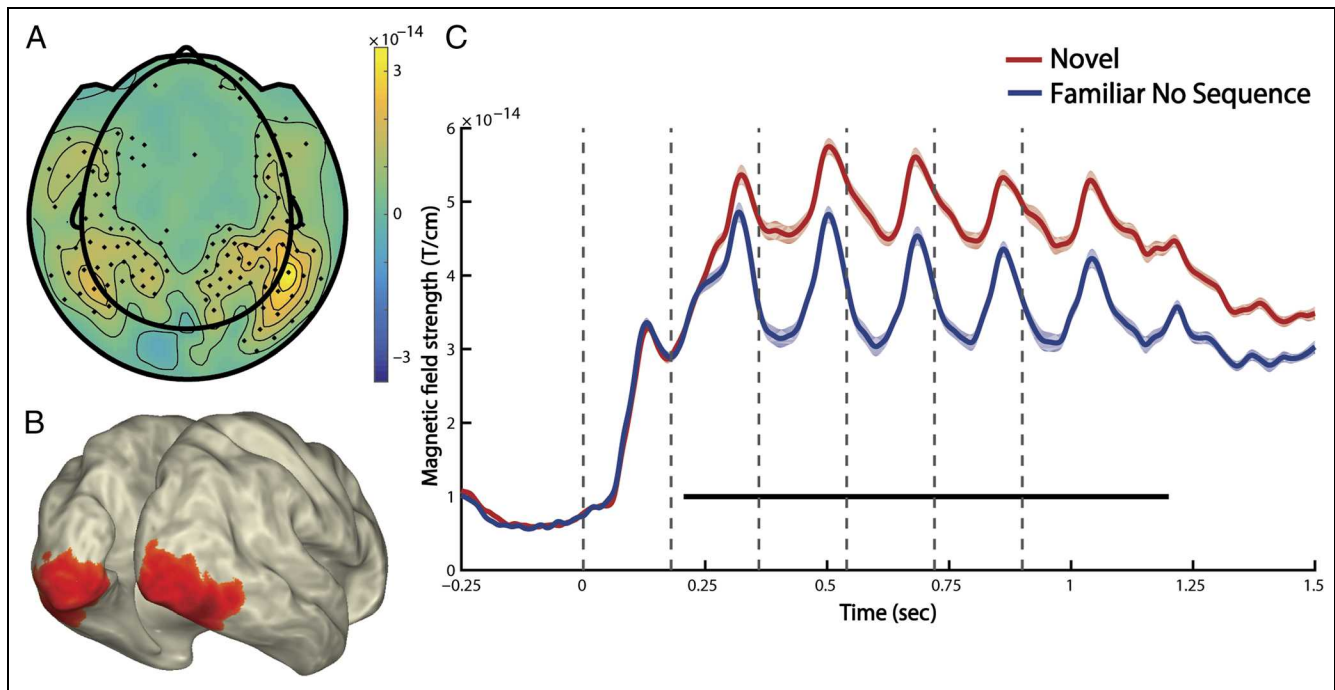
Figure 2. Effect of familiarity on evoked response. (A) Topography of the difference in amplitude between the Familiar No Sequence and Novel conditions (Novel − Familiar). Black asterisks mark sensors that contribute to the significant cluster for at least half of the time period from 200 to 1200 msec. (B) Source reconstruction of Familiar versus Novel. Activity was averaged over the time period of 200–1200 msec and interpolated onto a cortical surface. Plotted activity was thresholded at 80% of peak value for illustration purposes. (C) Activity over time for the Familiar No Sequence (blue) and Novel (red) conditions. Activity was averaged over sensors highlighted in A. Shaded areas are error bars illustrating within-subject *SEM* for the Familiar (light blue) and Novel (light red) conditions. Horizontal black bar at the bottom shows that at least one of the selected sensors contributes to the significant cluster at this time point. Dotted vertical lines denote the onset of each image.

Table 1. MNI Coordinates of Peak Source Locations for the Respective Comparisons

Amplitude Modulation			
	x	*y*	*z*
Novel vs. Familiar No Sequence Comparison			
Medial peak	3	−90	−26
Right occipital gyrus peak	46	−81	−18
Left occipital gyrus peak	−30	−87	−19
Unexpected vs. Expected Comparison			
Right occipital gyrus peak	39	−72	−14
Left occipital gyrus peak	−36	−77	−14

Dynamic Range Modulation			
	x	*y*	*z*
Novel vs. Familiar No Sequence Comparison			
Medial occipital peak	−8	−108	−8
Unexpected vs. Expected Comparison			
Right occipital peak	44	−84	−10

Reported values are mean values across participants in mm.

Unexpected Items Lead to Enhanced Activity in LOC

To examine the difference in amplitude between expected and unexpected items when familiarity was held constant, we compared the Familiar Expected Sequence and Familiar Unexpected Sequence conditions, both of which consisted of familiar images. A significant difference ($p = .008$) across conditions was observed, which was driven by the cluster of sensors shown in Figure 3A from approximately 500 msec until 900 msec. The black asterisks in the figure denote sensors that contribute to this cluster for at least half of the time period from 500 to 900 msec. The time course for the sensors contributing to the cluster is plotted in Figure 3C; the black line at the bottom shows that at least one of the selected sensors contributes to a significant difference at that time point. Evidently, unexpected items lead to significantly more activity than expected ones. Source analysis demonstrated that the difference between the Familiar Expected Sequence and Familiar Unexpected Sequence could be localized to the right inferior occipital gyrus and to a lesser degree to the left inferior occipital gyrus (Figure 3B; Table 1), corresponding to right and left LOC.

Moreover, we asked whether our data provided support for surprise enhancement or expectation suppression. To this end, we compared Familiar No Sequence items (i.e., where no sequence was learned) to both Familiar Unexpected
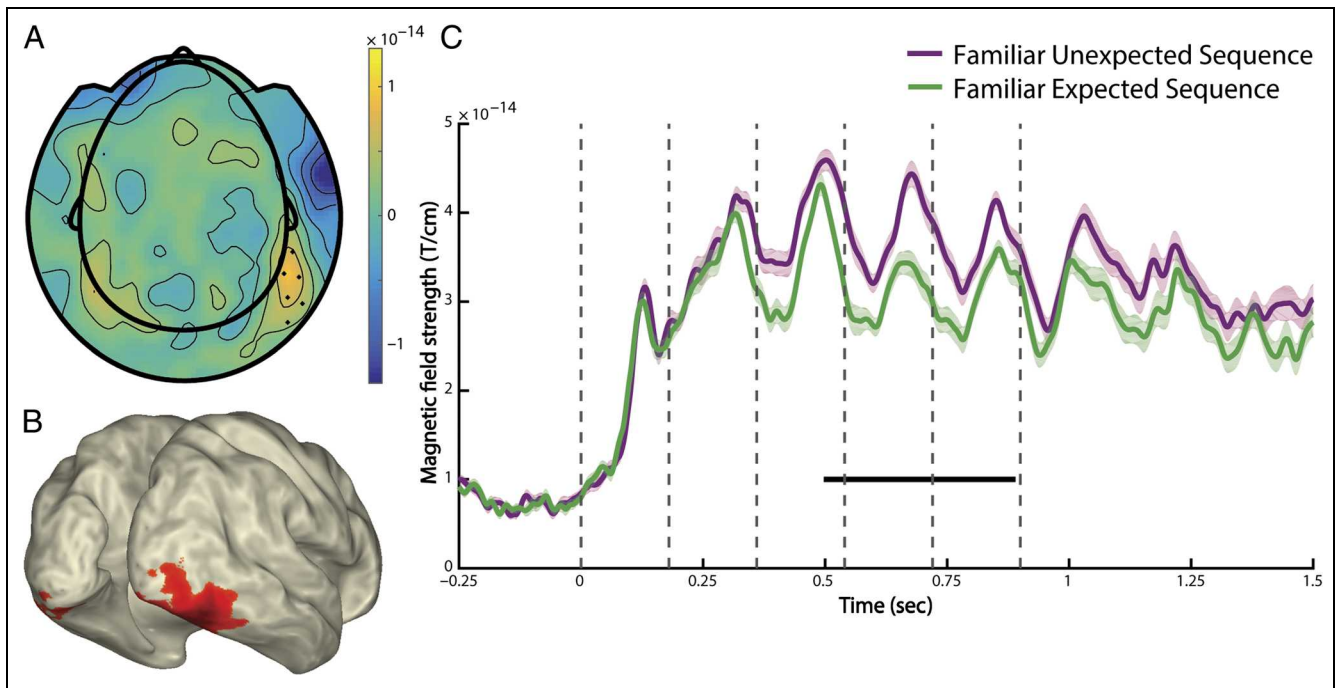
Figure 3. Expectation effect on amplitude. (A) Topography of the difference in amplitude between the Familiar Expected Sequence and Familiar Unexpected Sequence conditions (Unexpected − Expected). Black asterisks mark sensors that contribute to the significant cluster for at least half of the time period from 500 to 900 msec. (B) Source reconstruction of Expected versus Unexpected. Activity was averaged over the time period of 500–900 msec and interpolated onto a cortical surface. Plotted activity was thresholded at 80% of peak value for illustration purposes. (C) Activity over time for the Expected (green) and Unexpected (purple) conditions. Activity was averaged over sensors highlighted in A. Shaded areas are error bars illustrating within-subject *SEM* for the Expected (light green) and Unexpected (light purple) conditions. Horizontal black bar at the bottom shows that at least one of the selected sensors contributes to the significant cluster at this time point. Dotted vertical lines denote the onset of each image.

Sequence and Familiar Expected Sequence items, respectively, because the former contrast illustrates the effect of a violated expectation and the latter demonstrates the effect of a confirmed expectation (see Figure 4 for the time courses for all conditions). Familiar Unexpected Sequence trials

showed stronger activity than Familiar No Sequence trials ($t_{28} = -3.01, p = .006$), demonstrating an activity increase for violated expectations. Familiar Expected Sequence trials, on the other hand, did not show a robust difference to Familiar No Sequence trials ($t_{28} = -1.53, p = .14$).

Figure 4. Surprise enhancement and expectation suppression. Activity over time for Familiar Unexpected Sequence (purple) where the expectation was violated, Familiar Expected Sequence (green) where the expectation was confirmed, Familiar No Sequence (orange) where no sequence was learned, and Novel (blue). Shaded areas are error bars illustrating within-subject *SEM*. Dotted vertical lines denote the onset of each image. Activity was averaged over sensors highlighted in Figure 3A. A number of trials equal to the number of trials in the (Un)Expected condition is randomly selected from the

Novel and Familiar No Sequence conditions. The data provide evidence for surprise enhancement (Familiar Unexpected Sequence vs. Familiar No Sequence) but offer no statistical support for expectation suppression (Familiar Expected Sequence vs. Familiar No Sequence).
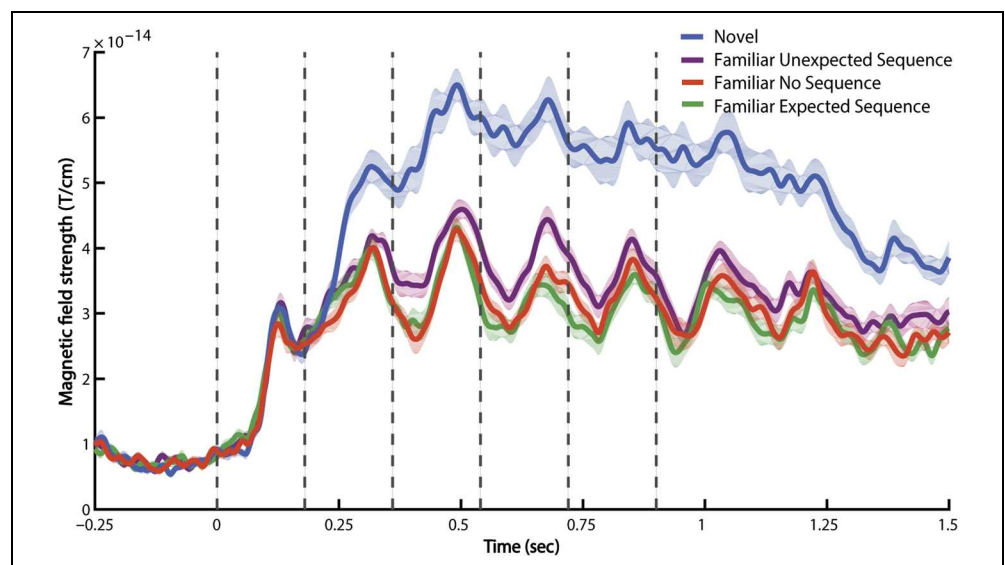
We also assessed whether the expectation effect changes over the course of the experiment. For this purpose, we divided the data into the first half and the second half of the experiment, calculated the expectation effect for each half, and compared them. There was no significant difference ($p = .39$) between the magnitudes of the expectation effects in the first half of the experiment compared with the second half.

Familiarity Leads to Higher Dynamic Range of the Sensory Response

Sharper response dynamics include a truncated sensory response, which leads to higher dynamic range (peak-to-trough excursion) of the response (Meyer et al., 2014). The dynamic range of the signal can be approximated by the power at the driving frequency. To investigate the difference in power at the stimulus frequency, we compared the Novel and Familiar No Sequence conditions. A significant difference ($p = .016$) emerged for the driving frequency of 5.6 Hz in the cluster of sensors shown in Figure 5A. The black asterisks in the figure denote sensors that contribute to the significant cluster, and the power spectrum for these sensors is plotted in Figure 5C. Familiar items led to significantly more power at the stimulus frequency of 5.6 Hz than novel items. Source reconstruction revealed that the largest difference in power between the familiar and novel conditions was in a medial posterior

location belonging to early occipital cortex (Figure 5B; Table 1).

Expectation Does Not Lead to Significantly Higher Dynamic Range of the Sensory Response

To examine whether expectation likewise led to a larger dynamic range of the neural response, we compared the difference in power at the stimulus frequency between the Familiar Unexpected Sequence and Familiar Expected Sequence conditions. There was no significant difference ($p = .700$) for the driving frequency of 5.6 Hz, as shown in Figure 6A. The power spectrum for all sensors is plotted in Figure 6C. It suggests that expected items may lead to more power at the stimulus frequency of 5.6 Hz than unexpected items, but this difference was not significant. Source analysis demonstrated that the largest difference in power between the Familiar Expected Sequence and Familiar Unexpected Sequence conditions was in early visual areas (Figure 6B; Table 1), but it is difficult to interpret this outcome because the effect is not statistically significant.

Explicit Knowledge of Sequence Immediately after Training Does Not Modulate the Expectation Effect

We wanted to determine whether the extent to which participants could report the order of the stimuli influenced
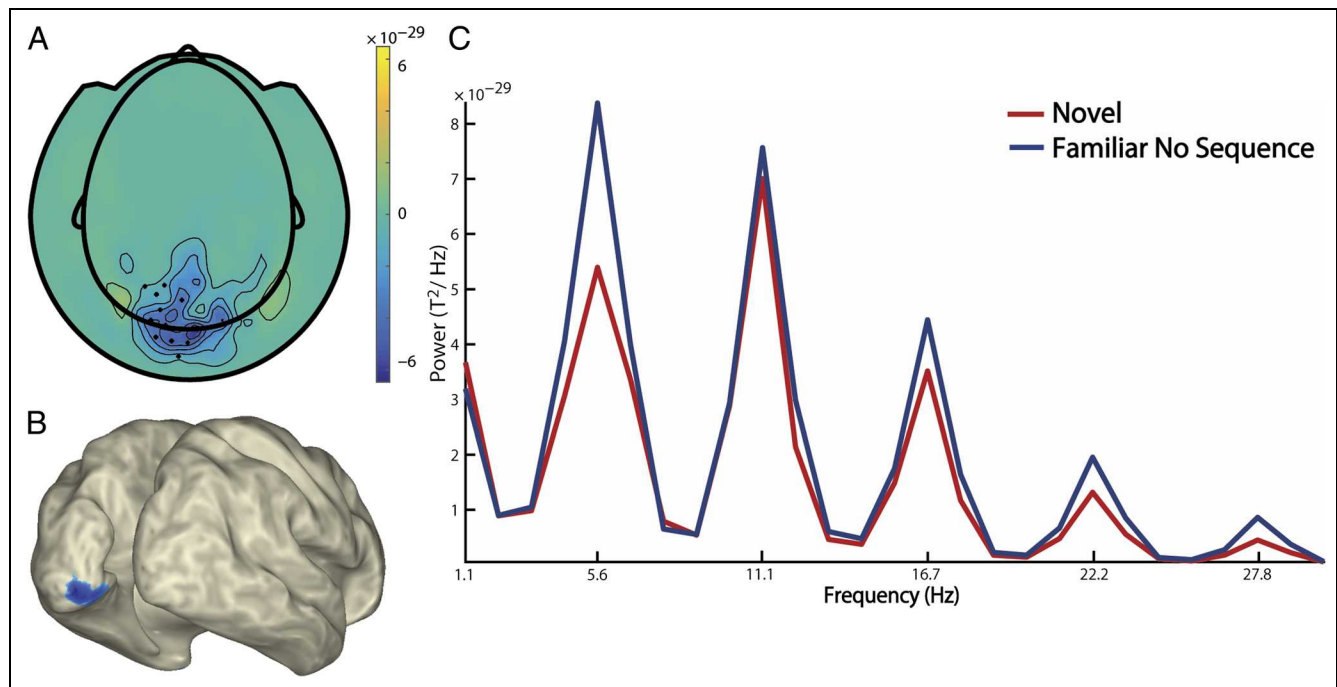

Figure 5. Effect of familiarity on power at the driving frequency. (A) Topography of the difference in power between the Familiar No Sequence and Novel conditions (Novel − Familiar). Black asterisks mark sensors that contribute to the cluster that showed a significant difference in power between conditions at the stimulus presentation frequency (5.6 Hz). (B) Source reconstruction of Familiar versus Novel. Activity corresponding to the frequency of 5.6 Hz was interpolated onto a cortical surface. Plotted activity was thresholded at 50% of peak value for illustration purposes. (C) Power spectrum of Familiar No Sequence (blue) and Novel (red). Activity was averaged over sensors highlighted in A.
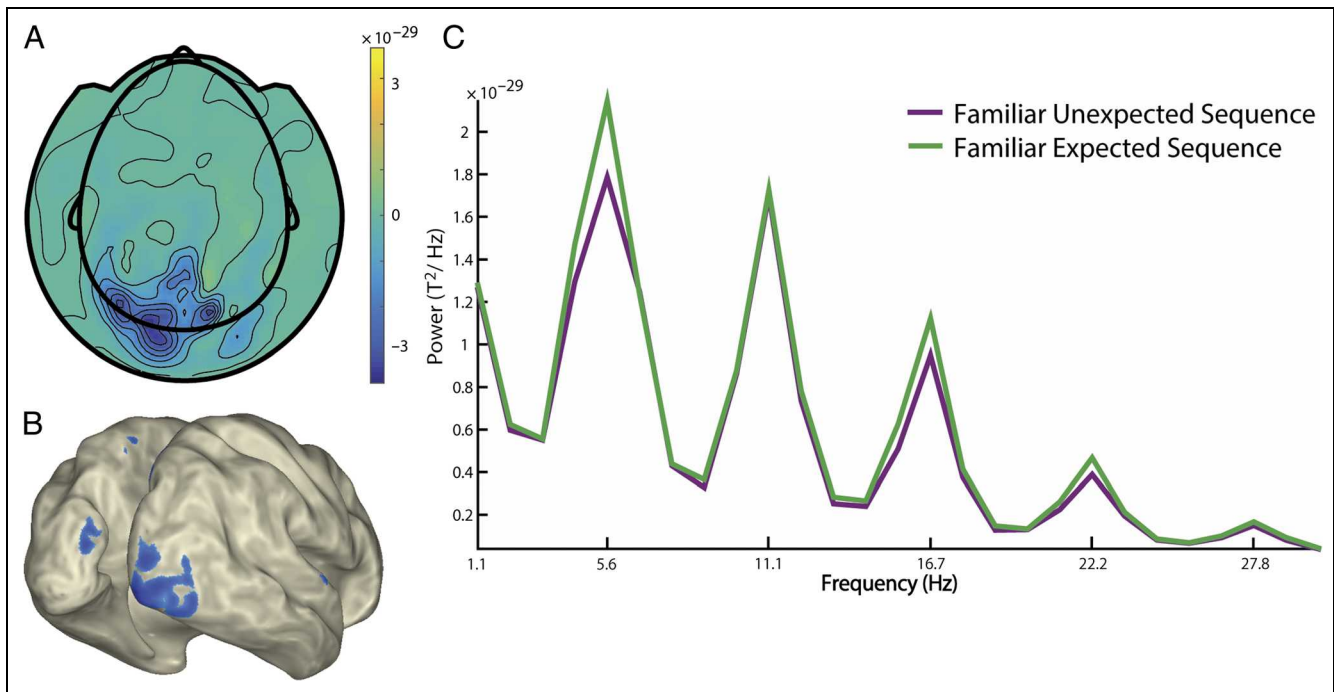
Figure 6. Effect of expectation on power at the driving frequency. (A) Topography of the difference in power between the Expected and Unexpected conditions (Unexpected − Expected). There was no significant difference between conditions at the stimulus presentation frequency (5.6 Hz). (B) Source reconstruction of Expected versus Unexpected. Activity corresponding to the frequency of 5.6 Hz was interpolated onto a cortical surface. Plotted activity was thresholded at 50% of peak value for illustration purposes. (C) Power spectrum of Expected (green) and Unexpected (purple). Activity was averaged over all sensors.

the expectation effect. Thus, we compared the expectation effect in the group which performed above chance on the sequence identification task to that in the group performing below chance. Although expectation modulated the amplitude of MEG activity in each group separately ($t_{13} = -3.28$, $p = .006$, for the above-chance group and $t_{13} = -2.83$, $p = .01$, for the below-chance group), this modulation did not differ significantly between the two groups ($t_{26} = -0.27$, $p = .79$). In addition, the Spearman's rank correlation showed that there was no significant correlation between participants' performance on the sequence learning task and the amplitude of the expectation modulation (rho $= -0.08$, $p = .68$). These results suggest that the effect of expectation on the magnitude of the neural response is not related to participants' explicit knowledge of the stimulus structure.

DISCUSSION

The sensory response to a stimulus can be modulated by whether a stimulus has been seen before (i.e., stimulus familiarity), as well as by whether a stimulus is expected or unexpected in the current context (i.e., stimulus expectation). In this study, we manipulated stimulus familiarity and expectation separately, which allowed us to examine the effects of familiarity and expectation on brain activity in the visual system within the same experimental paradigm, using MEG. We found that familiar images elicited markedly less neural activity than novel images in early

visual and object-selective LOC. Similarly, expected images were also associated with reduced neural activity compared with unexpected images in LOC. The independent manipulation of familiarity and expectation in our study allows us to conclude that these distinct types of sensory knowledge jointly modulate the amount of neural resources dedicated to object processing in the visual ventral stream.

Familiar items, when compared with novel items, were also associated with a temporally truncated (i.e., temporally sharpened) sensory response. In the context of our paradigm, in which stimuli rapidly followed each other, this led to an increased dynamic range of the signal. This was visible as increased power in the stimulus frequency, which was most prominent in early visual cortex. A similar trend, albeit nonsignificant, was present for expected versus unexpected items. The sharper response dynamics for familiar than novel stimuli we observed are in accordance with earlier findings by Meyer et al. (2014), although we observe the strongest contribution in early, rather than later, visual regions. Although an increase in attention can also lead to an increase of power in the stimulus frequency for visually entrained stimuli (Ding, Sperling, & Srinivasan, 2006), it is unlikely that this underlies the sharpened response for familiar items for two reasons. First, given that participants' task was to detect ducky stimuli, both familiar and novel items were equally relevant to the observer, precluding the need for stronger attentional engagement by the familiar items. Second, if anything,

familiarity would rather be associated with a reduction in attention, given that novel items are more salient and have stronger capacity to attract attention (Escera, Alho, Winkler, & Näätänen, 1998). Temporal sharpening (indexed by dynamic range) is a mechanism that has not been investigated extensively in humans yet, and this study suggests that it underlies familiarity and possibly expectation in the human visual cortex.

The experimental effects of familiarity and expectation had distinct time courses. The reduction in activity for familiar (compared with novel) items was apparent already 200 msec after the onset of the visual sequence, whereas the activity reduction for expected items only was visible after 500 msec. This is likely due to the fact that familiarity is already defined for the first image—which may have been seen before (familiar) or not (novel). With respect to expectation, the first image of a sequence could never be predicted; only after observing the first image, a prediction could be made about the subsequent images. This may explain why expectation effects are visible only later in the visual sequence. Apart from this, it is conceivable that activity modulation due to familiarity is the consequence of local changes in LOC (Vogels, 2016; Kaliukhovich & Vogels, 2011), whereas expectation suppression is implemented by feedback from regions higher up in the visual hierarchy (Friston, 2005) and potentially the hippocampus (Hindy, Ng, & Turk-Browne, 2016; Buckner, 2010), which send predictions down to LOC and V1. This may explain why the modulation by familiarity may be observed earlier than the modulation by expectation.

Our data suggest that activity modulations by expectation may mainly be caused by surprise enhancement rather than expectation suppression because unexpected images elicited significantly larger activity than items with no learned sequence, whereas expected images were not statistically different from images with no learned sequence in terms of neural activity. In line with this, Kaposvari et al. (2016) found that expectation suppression was of smaller magnitude than surprise enhancement, which opens the possibility that expectation suppression may be a weaker signal, and therefore, we did not observe it in our data. A recent study by Ramachandran, Meyer, and Olson (2017), on the other hand, found strong evidence for expectation suppression and only limited evidence for surprise enhancement. As argued by these authors, the apparent discrepancy between these studies may relate to the rate of presentation of the visual stimuli. Rate of presentation in both our study and the study by Kaposvari et al. (2016) was markedly faster than in the study by Ramachandran et al. (2017), potentially resulting in a weaker phasic response to the image and thereby less potential for expectation suppression.

Interestingly, we observed an effect of expectation although many participants could not explicitly report the sequence for the predictable images. During debriefing, participants reported that they did not notice any specific order for the images. The behavioral assessment of

sequence knowledge also showed that participants' performance was, on average, near chance level when they were shown an image and had to report which image should follow. Nevertheless, the neural response showed a distinct difference between expected and unexpected conditions, suggesting that the neural effects of expectation are likely due to implicit predictions that occur outside the awareness of the observer. This is in line with previous studies showing that subjects can learn transitions without becoming aware of them (Alamia et al., 2016; Clark & Squire, 1998; Reber, 1967).

In our data, increased dynamic range was most prominent in early visual cortex, whereas reduced amplitude was present most strongly in LOC. This may be related to the fact that these two brain areas have different temporal dynamics. In early visual cortex, visual input changes rapidly, in which case a sharpened response may be particularly useful for efficient coding. The temporal dynamics of LOC are more stable over time, making it possible for LOC to sustain an amplitude difference over time. Perhaps the sharpened response from early visual cortex, once integrated by areas higher along the visual stream, results in a reduced amplitude in LOC.

Although we isolated neural effects of familiarity and expectation by different manipulations, usually these concepts are heavily intertwined. In our experiment, the first image of a familiar stream may be much less surprising than a novel one, given that there were 12 familiar images and more than 2000 novel ones. In this sense, familiarity and predictability are indeed partly inseparable constructs. However, there is mounting evidence that, when keeping familiarity constant, the sensory response is modulated by whether a particular image is expected to occur at a particular moment in time (e.g., Meyer & Olson, 2011). Our paradigm allowed us to conclude that, when we hold familiarity constant and manipulate predictability (or expectation), we observe a highly similar modulation of the sensory response in terms of spatiotemporal profile. This leads us to conclude that these conceptually somewhat different forms of prior may be implemented similarly in the brain.

To conclude, this study demonstrates that familiarity and expectation jointly influence how much neural resources are dedicated to object processing in visual cortex. Familiar and expected inputs have similar effects on the brain signal because they reduce the amplitude of the response and increase its dynamic range. Notably, amplitude change was linked to LOC, whereas temporal sharpening was most prominent in early visual cortex, indicating that these two mechanisms may have different neural sources and underpinnings. These findings extend our understanding of how familiarity and expectation affect sensory processing in the human brain.

REFERENCES

Alamia, A., Orban de Xivry, J.-J., San Anton, E., Olivier, E., Cleeremans, A., & Zenon, A. (2016). Unconscious associative learning with conscious cues. *Neuroscience of Consciousness, 2016,* 1–10.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Müller, K.-R. (2011). Single-trial analysis and classification of ERP components—A tutorial. *Neuroimage, 56,* 814–825.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436.

Buckner, R. L. (2010). The role of the hippocampus in prediction and imagination. *Annual Review of Psychology, 61,* 27–48.

Clark, R. E., & Squire, L. R. (1998). Classical conditioning and brain systems: The role of awareness. *Science, 280,* 77–81.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1,* 42–45.

Davis, B., & Hasson, U. (2018). Predictability of what or where reduces brain activity, but a bottleneck occurs when both are predictable. *Neuroimage, 167,* 224–236.

Ding, J., Sperling, G., & Srinivasan, R. (2006). Attentional modulation of SSVEP power depends on the network tagged by the flicker frequency. *Cerebral Cortex, 16,* 1016–1029.

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *Journal of Neuroscience, 30,* 16601–16608.

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of Cognitive Neuroscience, 10,* 590–604.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 360,* 815–836.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences, 10,* 14–23.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience, 3,* 837–843.

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience, 19,* 665–667.

Kaliukhovich, D. A., & Vogels, R. (2011). Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex. *Cerebral Cortex, 21,* 1547–1558.

Kaposvari, P., Kumar, S., & Vogels, R. (2016). Statistical learning signals in macaque inferior temporal cortex. *Cerebral Cortex, 28,* 250–266.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A, 20,* 1434–1448.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164,* 177–190.

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences, U.S.A., 108,* 19401–19406.

Meyer, T., Walker, C., Cho, R. Y., & Olson, C. R. (2014). Image familiarization sharpens response dynamics of neurons in inferotemporal cortex. *Nature Neuroscience, 17,* 1388–1394.

Miller, E. K., Li, L., & Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. *Science, 254,* 1377–1379.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4,* 61–64.

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,* 1–9.

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Research, 1301,* 89–99.

Ramachandran, S., Meyer, T., & Olson, C. R. (2017). Prediction suppression and surprise enhancement in monkey inferotemporal cortex. *Journal of Neurophysiology, 118,* 374–382.

Rao, R. R. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2,* 79–87.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6,* 855–863.

Schoffelen, J.-M., Oostenveld, R., & Fries, P. (2008). Imaging the human motor system's beta-band synchronization during isometric contraction. *Neuroimage, 41,* 437–447.

St. John-Saaltink, E., Utzerath, C., Kok, P., Lau, H. C., & de Lange, F. P. (2015). Expectation suppression in early visual cortex depends on task set. *PLoS One, 10,* e0131172.

Stolk, A., Todorovic, A., Schoffelen, J.-M., & Oostenveld, R. (2013). Online and offline tools for head movement compensation in MEG. *Neuroimage, 68,* 39–48.

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience, 15,* 745–756.

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience, 30,* 11177–11187.

van Veen, B. D., van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering, 44,* 867–880.

Vogels, R. (2016). Sources of adaptation of inferior temporal cortical responses. *Cortex, 80,* 185–195.