Behavioral/Cognitive

Associative Prediction of Visual Shape in the Hippocampus

Peter Kok[1,2] and Nicholas B. Turk-Browne[1,2]

[1]Princeton Neuroscience Institute and Department of Psychology, Princeton University, Princeton, New Jersey 08544 and [2]Department of Psychology, Yale University, New Haven, Connecticut 06520

Perception can be cast as a process of inference, in which bottom-up signals are combined with top-down predictions in sensory systems. In line with this, neural activity in sensory cortex is strongly modulated by prior expectations. Such top-down predictions often arise from cross-modal associations, such as when a sound (e.g., bell or bark) leads to an expectation of the visual appearance of the corresponding object (e.g., bicycle or dog). We hypothesized that the hippocampus, which rapidly learns arbitrary relationships between stimuli over space and time, may be involved in forming such associative predictions. We exposed male and female human participants to auditory cues predicting visual shapes, while measuring high-resolution fMRI signals in visual cortex and the hippocampus. Using multivariate reconstruction methods, we discovered a dissociation between these regions: representations in visual cortex were dominated by whichever shape was presented, whereas representations in the hippocampus reflected only which shape was predicted by the cue. The strength of hippocampal predictions correlated across participants with the amount of expectation-related facilitation in visual cortex. These findings help bridge the gap between memory and sensory systems in the human brain.

Key words: **expectation; hippocampal subfields; perceptual inference; predictive coding; shape perception**

Significance Statement

The way we perceive the world is to a great extent determined by our prior knowledge. Despite this intimate link between perception and memory, these two aspects of cognition have mostly been studied in isolation. Here we investigate their interaction by asking how memory systems that encode and retrieve associations can inform perception. We find that upon hearing a familiar auditory cue, the hippocampus represents visual information that had previously co-occurred with the cue, even when this expectation differs from what is currently visible. Furthermore, the strength of this hippocampal expectation correlates with facilitation of perceptual processing in visual cortex. These findings help bridge the gap between memory and sensory systems in the human brain.

Introduction

Neural activity in sensory cortex can be strongly modulated by prior expectations (Summerfield et al., 2008; den Ouden et al., 2009; Alink et al., 2010; Meyer and Olson, 2011; Todorovic et al., 2011; Wacongne et al., 2011; Kok et al., 2012, 2013; Bell et al., 2016; Kaposvari et al., 2016). Most theories of the neural mechanisms underlying such phenomena focus on low-level, highly ingrained predictions, such as surround suppression or filling-in of contours (Lee and Nguyen, 2001; Spratling, 2010; Kok and De Lange, 2014), which may be represented by downstream areas within local brain circuits (Rao and Ballard, 1999; Spratling, 2010). However, it is unclear how to extend these proposals to more complex, learned predictions. Consider cross-modal predictions, for example, such as when an auditory stimulus (e.g., a bell or bark) leads to an expectation of the visual appearance of the corresponding object (e.g., a bicycle or dog). Such associations, especially when learned recently, cannot readily be encoded within sensory systems, as visual cortex does not have direct access to the features of auditory stimuli nor is it able to rapidly bind these features. Such predictions may instead depend on a higher-order brain region that can rapidly learn multisensory associations.

Based on this, we hypothesized that the hippocampus plays a role in such predictions. First, the hippocampus is known to be involved in learning arbitrary relationships between stimuli (Cohen and Eichenbaum, 1993; Davachi, 2006; Turk-Browne et al., 2009; Henke, 2010; Hsieh et al., 2014; Garvert et al., 2017), particularly over space and time (Solomon et al., 1986; Wallenstein et al., 1998; Staresina and Davachi, 2009). In fact, learning of such relationships is strongly impaired when the hippocampus is dam-

Received Jan. 21, 2018; revised April 29, 2018; accepted June 20, 2018.

Author contributions: P.K. and N.B.T.-B. designed research; P.K. performed research; P.K. analyzed data; P.K. and N.B.T.-B. wrote the paper.

This work was supported by NWO Rubicon Grant 446-15-004 to P.K. and NIH Grants R01 EY021755 and R01 MH069456 to N.B.T.-B. We thank Nicholas C. Hindy and Mariam Aly for help with hippocampal segmentations.

The authors declare no competing financial interests.

Correspondence should be addressed to Peter Kok, Department of Psychology, Yale University, 2 Hillhouse Avenue, New Haven, CT 06520. E-mail: peter.kok@yale.edu.

DOI:10.1523/JNEUROSCI.0163-18.2018

 0270-6474/18/386888-12$15.00/0

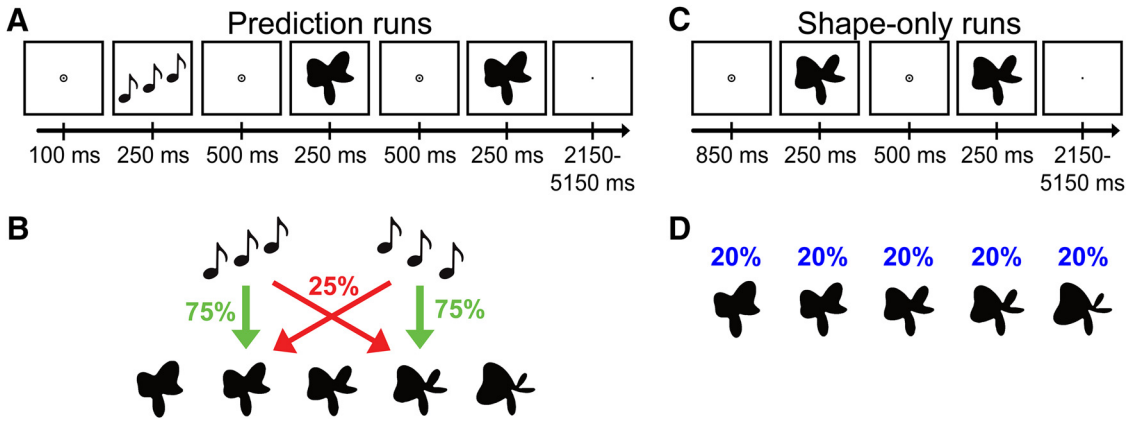

Figure 1. Experimental paradigm. ***A***, During prediction runs, an auditory cue preceded the presentation of two consecutive shape stimuli. On each trial, the second shape was either identical to the first or slightly warped with respect to the first along an orthogonal dimension, and participants' task was to report whether the two shapes were the same or different. ***B***, The auditory cue (ascending vs descending tones) predicted whether the first shape on that trial would be shape 2 or shape 4 (of 5 shapes). The cue was valid on 75% of trials, whereas in the other 25% of (invalid) trials the unpredicted shape was presented. ***C***, During shape-only runs, no auditory cues were presented. As in the prediction runs, two shapes were presented on each trial, and participants' task was to report same or different. ***D***, All five shapes appeared with equal (20%) likelihood on trials of the shape-only runs.

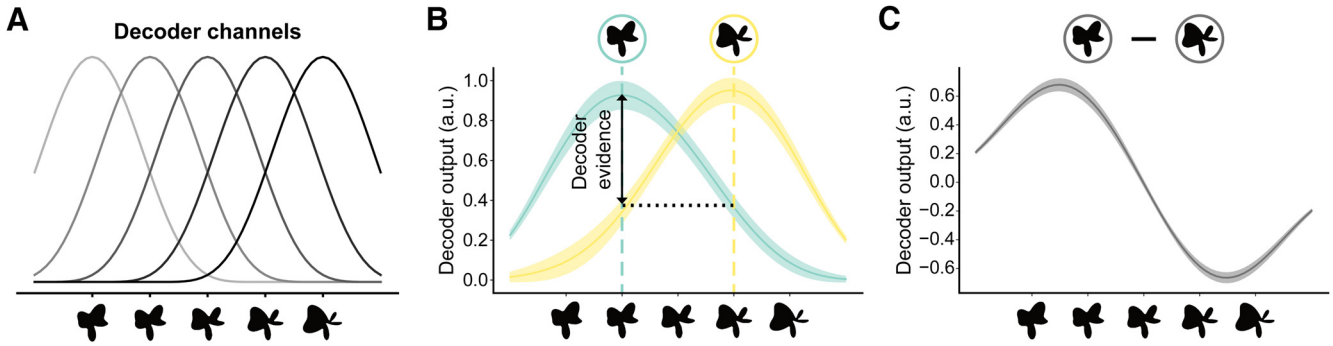

Figure 2. Illustration of the decoding method. ***A***, We used a forward modeling approach to reconstruct shapes from the pattern of BOLD activity. Shape selectivity was characterized by five hypothetical channels, each with an idealized shape tuning curve. BOLD patterns obtained from the shape-only runs were used to estimate the weights on the five hypothetical channels separately for each voxel, using linear regression. ***B***, Using these weights, the second stage of the analysis reconstructed the channel outputs associated with the pattern of activity across voxels evoked by the prediction runs (only shapes 2 and 4 were used in these runs). Channel outputs were converted to a weighted average of the five basis functions, resulting in neural shape tuning curves. Decoding performance was quantified by subtracting the amplitude of the shape tuning curve at the presented shape (e.g., shape 2) from the amplitude at the nonpresented shape (e.g., shape 4). ***C***, Finally, we collapsed across the presented shapes by subtracting the shape tuning curve for shape 4 from that for shape 2, thereby removing any nonshape-specific BOLD signals. Shaded regions in ***B*** and ***C*** indicate SEM.

aged (Sutherland et al., 1989; Chun and Phelps, 1999; Hannula et al., 2006; Konkel et al., 2008; Schapiro et al., 2014). Second, the hippocampus has bidirectional connections with sensory cortices of all modalities (Lavenex and Amaral, 2000; Eichenbaum et al., 2007; Henke, 2010) and has, for this reason, even been considered the top of sensory hierarchies (Felleman and Van Essen, 1991). Third, one of the main computational functions of the hippocampus is to retrieve associated items from memory based on partial information, a process known as pattern completion (Treves and Rolls, 1994; McClelland et al., 1995; Henke, 2010). This function has been mostly considered in the context of recall from episodic memory but is also ideally suited for retrieving predictions based on contextual cues (McClelland et al., 1995; Chun and Phelps, 1999; Eichenbaum and Fortin, 2009; Schapiro et al., 2012; Davachi and DuBrow, 2015; Hindy et al., 2016). Pattern completion is thought to be subserved by the CA3 subfield of the hippocampus, because of its strong recurrent, autoassociative connections (Treves and Rolls, 1994; Henke, 2010; Schapiro et al., 2017), from whence the retrieved pattern is sent to CA1, where it may be compared with actual sensory inputs (Lisman and Grace, 2005; Chen et al., 2011; Duncan et al., 2012). These computational mechanisms may allow the hippocampus to represent expected events, based on predictive cues in the environment, rather than being dominated by current inputs as in sensory cortices.

To investigate the involvement of the hippocampus in cross-modal predictions, we exposed human participants to auditory tones preceding the appearance of particular visual shapes (Fig. 1), while measuring blood oxygenation level-dependent (BOLD) signals in both visual cortex and the hippocampus with high-resolution functional magnetic resonance imaging (fMRI). Using multivariate pattern analysis and an inverted encoding model (Fig. 2), we reconstructed which shape was represented in these brain systems on trials in which the tones validly versus invalidly predicted what appeared. We hypothesized that the hippocampus would represent the shape expected based on the tones regardless of which shape actually appeared. In contrast, whereas visual cortex may be modulated by expectation, its representation should be dominated by the shape presented on screen.

Materials and Methods

Participants. Twenty-five healthy individuals participated in the experiment. All participants were right-handed, were MR compatible, and had normal or corrected-to-normal vision. Participants provided informed

consent to a protocol approved by the Princeton University Institutional Review Board and were compensated ($20 per hour). One participant was excluded from analysis because they moved their head between runs such that large parts of the occipital lobe were no longer inside the field of view. The final sample consisted of 24 participants (15 female; age, 23 ± 3, mean ± SD).

Stimuli. Visual stimuli were generated using MATLAB (Mathworks; RRID:SCR_001622) and the Psychophysics Toolbox (Brainard, 1997; RRID:SCR_002881). In the MR scanner, the stimuli were displayed on a rear-projection screen using a projector (1024 × 768 resolution, 60 Hz refresh rate) against a uniform gray background. Participants viewed the visual display through a mirror that was mounted on the head coil. The visual stimuli consisted of complex shapes defined by radial frequency components (RFCs) (Zahn and Roskies, 1972; Op de Beeck et al., 2001; Drucker and Aguirre, 2009; Fig. 1*A*). The contours of the stimuli were defined by seven RFCs, based on a subset of the stimuli used by Op de Beeck et al. (2001; see their Fig. 1*a*). A one-dimensional shape space was created by varying the amplitude of three of the seven RFCs. Specifically, the amplitudes of the 1.11, 1.54, and 4.94 Hz components increased together, ranging from 0 to 36 (first two components) and from 15.58 to 33.58 (third component). Note that we chose to vary three RFCs simultaneously, rather than one, to increase the perceptual (and neural) discriminability of the shapes.

To map out this shape space perceptually, we generated 13 shapes that spanned the continuum, with the amplitudes of the three modulated RFCs increasing with equal steps from the minimum to the maximum of the ranges defined above. Six participants categorized these shapes as one of the two extremes of the continuum (each shape presented 24 times). Psychometric curves were fit to these data, and we determined the points along the continuum (in terms of the amplitudes of the three modulated RFCs) that were judged as 10, 50, and 90% likely to be the extreme shape with maximal values. The five shapes we used in the fMRI experiment consisted of these three experimentally determined points in the space continuum, as well as the two extremes (Fig. 1*A*). The participants who took part in the fMRI experiment were exposed to the same perceptual categorization experiment after the fMRI session ended, and we determined that for each participant the chance of classifying a shape as the maximal extreme increased monotonically as a function of the amplitude of the three RFCs. During the fMRI experiment, the shapes were presented centered on fixation (color, black; size, 4.5°). Additionally, a fourth RFC (the 3.18 Hz component) was used to create slightly warped versions of the five shapes, to enable the same/different shape discrimination cover task (see below).

Auditory cues consisted of three pure tones (440, 554, and 659 Hz; 80 ms per tone; 5 ms intervals), presented in either ascending or descending pitch.

Experimental procedure. Each trial of the main experiment started with the presentation of a fixation bullseye (diameter, 0.7°). During the prediction runs, an auditory cue (ascending or descending tones, 250 ms) was presented 100 ms after onset of the trial (Fig. 1*A*). After a 500 ms delay, two consecutive shape stimuli were presented for 250 ms each, separated by a 500 ms blank screen (Fig. 1*A*). The auditory cue (ascending vs descending tones) predicted whether the first shape on that trial would be shape 2 or shape 4 (of five shapes; Fig. 1*B*). The cue was valid on 75% of trials, whereas in the other 25% of trials the unpredicted shape would be presented. For instance, an ascending auditory cue might be followed by shape 2 on 75% of trials and by shape 4 on the remaining 25% of trials. Participants were trained on the cue–shape associations during practice runs that took place immediately before the prediction runs, in the scanner. That is, before the first prediction run, participants performed a practice run, consisting of two blocks of 56 trials each (112 trials total, ~8 min), in which the auditory cue was 100% predictive of the identity of the first shape on that trial (e.g., ascending tones always followed by shape 2 and descending tones followed by shape 4). Halfway through the experiment, the contingencies between the auditory cues and the shapes were flipped (e.g., ascending tones now followed by shape 4 and descending tones by shape 2), and participants performed another practice run (112 trials, ~8 min) to learn the new contingencies. The order of the cue–shape mappings was counterbalanced across participants. This procedure served to equate the frequencies of all tones and shapes and their transitions and to ensure that any differences between valid and invalid trials could not be explained by stimulus differences. The two practice runs took place while anatomical scans (see below) were acquired, to make full use of scanner time. Note that learning of the associations was not explicitly assessed behaviorally, but rather relied on previous work using a highly similar task structure that resulted in rapid learning (Kok et al., 2012, 2014, 2017).

On each trial, the second shape was either identical to the first or slightly warped. This warp was achieved by modulating the amplitude of the 3.18 Hz RFC component defining the shape. This modulation could be either positive or negative (counterbalanced over conditions), and participants' task was to indicate whether the two shapes on a given trial were the same or different, using an MR-compatible button box. After the response interval ended (750 ms after disappearance of the second shape), the fixation bullseye was replaced by a single dot, signaling the end of the trial while still requiring participants to fixate. This task was designed to avoid a direct relationship between the perceptual prediction and the task response. Furthermore, by modulating one of the RFCs that was not used to define our one-dimensional shape space, we ensured that the shape change on which the task was performed was orthogonal to the changes that defined the shape space and thus orthogonal to the prediction cues. The size of the modulation was determined by an adaptive staircasing procedure (Watson and Pelli, 1983), updated after each trial, to make the task challenging (~75% correct). Separate staircases were run for trials containing valid and invalid cues, as well as for the shape-only runs, to equate task difficulty between conditions. All participants completed two runs of this task (128 trials per run).

In two additional runs, which were interleaved with the runs just described (in ABBA fashion, order counterbalanced over participants), the 25% invalid trials did not involve presentation of the unpredicted shape, but rather no shape stimuli were presented at all. These omission trials were included in an attempt to decode expected but omitted shapes from the BOLD response. However, no such effects were found. One potential explanation of this null finding may be that the omission of the shape triggered different cognitive processes than during valid and invalidly cues trials. For example, the absence of any shape is quite salient and surprising given the regularity of their appearance in the rest of the study. In addition, participants did not perform a task on the omission trials, eliminating the need for perceptual discrimination, decision processes, and response selection. However, there are other potential explanations as well, perhaps related to the nature of the shape stimuli. Specifically, it is striking that expected but omitted shapes could not be decoded from visual cortex, whereas expected but omitted gratings can be in a highly similar design (Kok et al., 2014). We are conducting additional studies to better understand the conditions under which omission trials do (Kok et al., 2014; Hindy et al., 2016) and do not (the current study) reveal expectations, and thus the omission data are not considered further in this study.

Finally, participants completed two shape-only runs (120 trials per run), in which no auditory cues were presented. As in the prediction runs, the start of each trial was signaled by the onset of the fixation bullseye, and the stimulus onset asynchrony (SOA) between this onset and the presentation of the first shape was 850 ms (Fig. 1*C*). On any given trial, one of the five shapes would be presented, with equal (20%) likelihood. As in the prediction runs, the first shape was followed by a second one that was either identical or slightly warped, and participants' task was to report same or different. These runs were designed to be as similar as possible to the prediction runs, save the absence of the auditory cues and the equal rates of presentation of all five shapes. The two shape-only runs flanked the runs containing the auditory cues, constituting the first run and sixth (last) run of the experiment.

The staircases were kept running throughout the experiment. They were initialized at a value determined during an initial practice session 1–3 d before the fMRI experiment (no auditory cues, 120 trials). After the initial practice run, the meaning of the auditory cues was explained, and participants practiced briefly with both cue–shape contingencies (valid trials only; 16 trials per contingency). Note that this practice session did not train participants on any particular cue–shape association, since both

associations were practiced equally often, but instead simply served to acquaint them with the structure of the trials in the fMRI session. The actual training of the cue–shape associations took place in the scanner (see above).

MRI acquisition. Structural and functional MRI data were collected on a 3T Siemens Prisma scanner with a 64-channel head coil. Functional images were acquired using a multiband echoplanar imaging sequence (TR, 1000 ms; TE, 32.6 ms; 60 transversal slices; voxel size, 1.5 × 1.5 × 1.5 mm; 55° flip angle; multiband factor, 6). This sequence produced a partial volume for each participant, parallel to the hippocampus and covering the majority of the temporal and occipital lobes. Anatomical images were acquired using a T1-weighted MPRAGE sequence, using a GeneRalized Autocalibrating Partial Parallel Acquisition (GRAPPA) acceleration factor of 3 (TR, 2300 ms; TE, 2.27 ms; voxel size, 1 × 1 × 1 mm; 192 transversal slices; 8° flip angle). Additionally, to enable hippocampal segmentation, two T2-weighted turbo spin-echo (TSE) images (TR, 11,390 ms; TE, 90 ms; voxel size, 0.44 × 0.44 × 1.5 mm; 54 coronal slices; perpendicular to the long axis of the hippocampus; distance factor, 20%; 150° flip angle) were acquired. To correct for susceptibility-induced distortions in the echoplanar images, a pair of spin-echo volumes was acquired in opposing phase-encode directions (anterior/posterior and posterior/anterior) with matching slice prescription, voxel size, field of view, bandwidth, and echo spacing (TR, 8000 ms; TE, 66 ms).

fMRI preprocessing. The images were preprocessed using FEAT 6 (FMRI Expert Analysis Tool), part of FSL 5 (http://fsl.fmrib.ox.ac.uk/fsl, Oxford Centre for Functional MRI of the Brain; RRID:SCR_002823; Jenkinson et al., 2012). Susceptibility-induced distortions were determined on the basis of the opposing spin-echo volume pairs using the FSL topup tool (Andersson et al., 2003). The resulting off-resonance field output was converted from hertz to radians per second and supplied to FEAT for B0 unwarping (see below). The first six volumes of each run were discarded to allow T1 equilibration. For each run, the remaining functional images were spatially realigned to correct for head motion, and simultaneously supplied to B0 unwarping and registered to the participants' structural T1 image, using boundary-based registration. The functional data were temporally high-pass filtered with a 128 s period cutoff; no spatial smoothing was applied. Finally, the two TSE images were averaged, and the resulting image was registered to the T1 image through FLIRT (FMRIB's Linear Image Registration Tool).

All analyses were performed in participants' native space. For the searchlight analyses (see below), each participant's output volumes were registered to the Montreal Neurological Institute (MNI) template to allow group-level statistics. This was achieved by applying the nonlinear registration parameters obtained from registering each participant's T1 image to the MNI template using AFNI's (RRID:SCR_005927) 3dQwarp (https://afni.nimh.nih.gov/pub/dist/doc/program_help/3dQwarp.html).

Regions of interest. Hippocampal subfields CA2–CA3–DG, CA1, and the subiculum were defined on the basis of the TSE and T1 images using the automatic segmentation of hippocampal subfields machine learning toolbox (Yushkevich et al., 2015) and a database of manual medial temporal lobe segmentations from a separate set of 51 participants (Aly and Turk-Browne, 2016a,b). Manual segmentations were based on anatomical landmarks used in prior studies (Duvernoy, 2005; Carr et al., 2010; Schapiro et al., 2012). Consistent with these studies, CA2, CA3, and DG were combined into a single region of interest (ROI) because these subfields are difficult to distinguish at our functional resolution (1.5 mm isotropic). TSE acquisition failed for one participant, and so their hippocampal ROIs were based on the T1 image alone. Results of the automated segmentation were inspected visually for each participant. The hippocampus ROI consisted of the union of the CA2–CA3–DG, CA1, and subiculum subfields.

In visual cortex, V1, V2, and lateral occipital (LO) cortex were automatically defined in each participant's T1-weighted anatomical scan with FreeSurfer (http://surfer.nmr.mgh.harvard.edu/; RRID:SCR_001847). Finally, putamen and caudate ROIs were obtained from FreeSurfer's subcortical segmentation, since these regions have been implicated in associative learning and prediction (Poldrack et al., 2001; den Ouden et al., 2009; Turk-Browne et al., 2009; Shohamy and Turk-Browne, 2013).

The visual cortex ROIs were restricted to the 500 most active voxels during the shape-only runs in each ROI, to ensure that we were measuring responses in the retinotopic locations corresponding to our visual stimuli. Since no clear retinotopic organization is present in the other ROIs, cross-validated feature selection was used instead (see below).

All ROIs were collapsed over the left and right hemispheres since we had no hypotheses regarding hemispheric differences.

fMRI data modeling. The functional data of each participant were modeled with general linear model, using FILM (FMRIB's Improved Linear Model), which included temporal autocorrelation correction and extended motion parameters (six standard parameters, plus their derivatives and their squares) as nuisance covariates. We specified regressors for the conditions of interest [shape-only runs, five shapes; prediction runs, two shapes × two prediction conditions (valid vs invalid)], by convolving a delta function at the onset of the first shape on each trial with a double-gamma hemodynamic response function (HRF). Additionally, we included the temporal derivative of each regressor to accommodate variability in the onset of the response (Friston et al., 1998).

To investigate the temporal evolution of shape representations in visual cortex, a finite impulse response (FIR) approach was used to estimate the BOLD signal evoked by each condition of interest in 20 1 s intervals. This allowed us to estimate the shape decoding signal in a time-resolved manner, by training the decoder on the FIR parameter estimates from the 4–7 s time bins in the shape-only runs (corresponding to the peak hemodynamic signal) and applying it to all time bins for the prediction runs. The amplitude and latency of this time-resolved decoding signal was quantified by fitting a double-gamma function and its temporal derivative.

Shape decoding. To probe neural shape representations, we used a forward modeling approach to reconstruct the shape from the pattern of BOLD activity in a given brain region (Brouwer and Heeger, 2009). This approach has proven successful in reconstructing continuous stimulus features, such as hue (Brouwer and Heeger, 2009), orientation (Brouwer and Heeger, 2011), and motion direction (Kok et al., 2013). In the current study, shape contour was constructed along a continuous dimension (see above), allowing the application of a forward model.

We characterized the shape selectivity of each voxel as a weighted sum of five hypothetical channels, each with an idealized shape tuning curve (or basis function). As in previous forward model implementations (Brouwer and Heeger, 2009, 2011; Kok et al., 2013), each basis function consisted of a half-wave-rectified sinusoid raised to the fifth power, and the five basis functions were spaced evenly, such that they were centered on the five points in shape space that constituted the five shapes presented in the experiment (Fig. 2*A*). As a result of this, a tuning curve with any possible shape preference (within the space defined here) could be expressed as a weighted sum of the five basis functions. Note that, unlike other stimulus features previously reconstructed using forward models, the shape space used here was not circular and therefore the channels did not wrap around.

In the first stage of the analysis, we used parameter estimates obtained from the two shape-only runs to estimate the weights on the five hypothetical channels separately for each voxel, using linear regression. Specifically, let k be the number of channels, m the number of voxels, and n the number of measurements (i.e., the five shapes). The matrix of estimated response amplitudes for the different shapes during the shape-only runs ($\mathbf{B_{so}}$, $m \times n$) was related to the matrix of hypothetical channel outputs ($\mathbf{C_{so}}$, $k \times n$) by a weight matrix ($\mathbf{W}$, $m \times k$): $\mathbf{B_{so}} = \mathbf{W}\mathbf{C_{so}}$.

The least-squares estimate of this weight matrix W was estimated using linear regression: $\hat{\mathbf{W}} = \mathbf{B_{so}}\mathbf{C_{so}}^{\mathrm{T}}(\mathbf{C_{so}}\mathbf{C_{so}}^{\mathrm{T}})^{-1}$.

These weights reflected the relative contribution of the five hypothetical channels in the forward model (each with their own shape selectivity) to the observed response amplitude of each voxel. Using these weights, the second stage of analysis reconstructed the channel outputs associated with the pattern of activity across voxels evoked by the stimuli in the main experiment ($\mathbf{B_{exp}}$), again using linear regression. This step transformed each vector of n voxel responses (parameter estimates per condition) into a vector of five (number of basis functions) channel responses. More specifically, the channel responses ($\mathbf{C_{exp}}$) associated with the responses in the main experiment ($\mathbf{B_{exp}}$) were estimated using the learned weights ($\mathbf{W}$): $\hat{\mathbf{C}}_{\mathbf{exp}} = (\hat{\mathbf{W}}^{\mathrm{T}}\hat{\mathbf{W}})^{-1}\hat{\mathbf{W}}^{\mathrm{T}}\hat{\mathbf{B}}_{\mathbf{exp}}$.

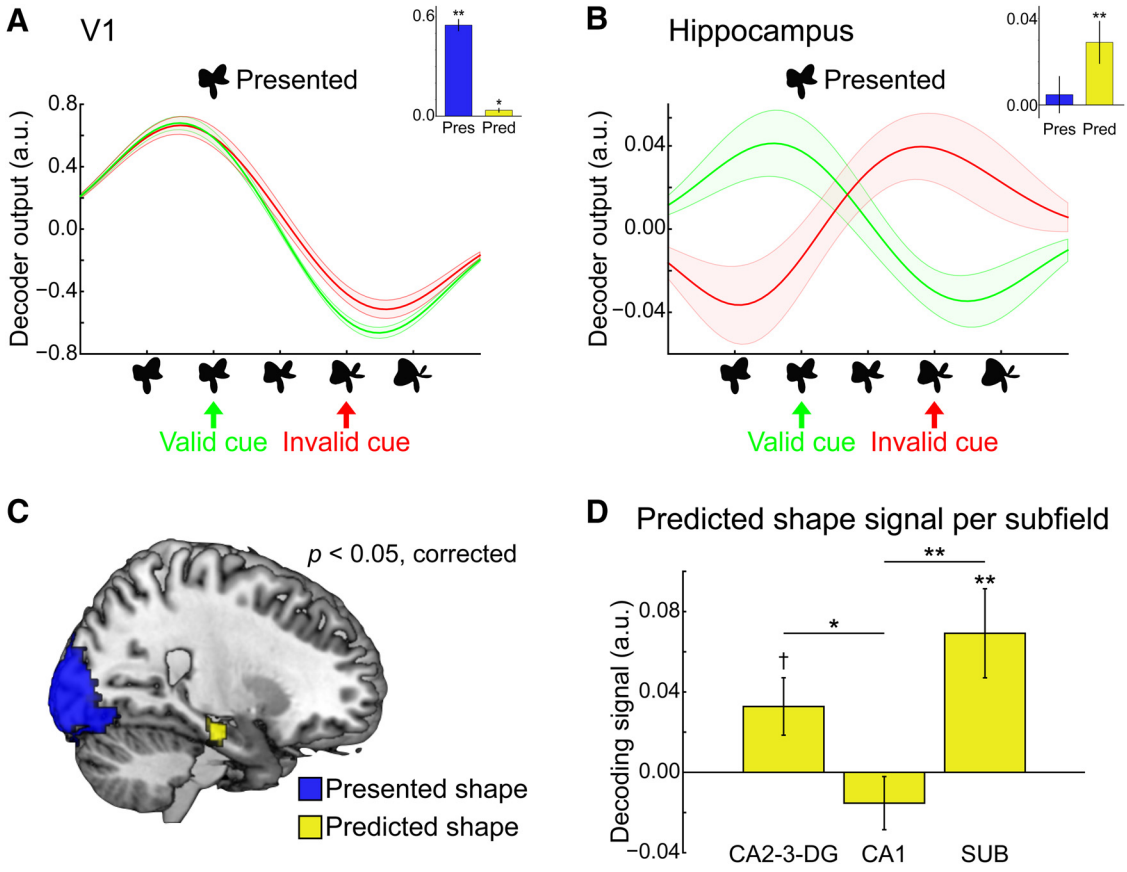

Figure 3. Shape representations in visual cortex and hippocampus. ***A***, Shape reconstructions from patterns of activity in V1, separately for validly (green) and invalidly (red) predicted shapes. Representations in visual cortex (V1, V2, LO; V1 plotted as representative region) were dominated by the presented shape, with modest modulation by the predictive cues in V1. The inset depicts quantified evidence for presented (Pres, blue) and predicted (Pred, yellow) shapes. ***B***, Shape reconstructions in the hippocampus were fully determined by the cued (predicted) shape, rather than the presented shape. The inset depicts quantified evidence for presented (Pres, blue) and predicted (Pred, yellow) shapes. ***C***, A searchlight analysis revealed evidence for the presented shape in the occipital lobe and for the predicted (but not presented) shape in the hippocampus. See Table 1 for full results. ***D***, Decoding of the predicted shapes across hippocampal subfields (SUB). $^{\dagger}p = 0.053$; $^{*}p < 0.05$; $^{**}p < 0.01$. Shaded regions and error bars indicate SEM.

These channel outputs were used to compute a weighted average of the five basis functions, reflecting a neural shape tuning curve (Fig. 2*B*). Note that, during the main experiment (i.e., the prediction runs), only shapes 2 and 4 were presented. Thus, four shape tuning curves were obtained for the prediction runs: two shapes by two prediction conditions (valid vs invalid). We collapsed across the presented shape by subtracting the shape tuning curve for shape 4 from that for shape 2, thereby subtracting out any nonshape-specific BOLD signals (Fig. 2*C*).

Decoding performance was quantified by subtracting the amplitude of the shape tuning curve at the presented shape (e.g., shape 2) from the amplitude at the nonpresented shape (shape 4). Collapsing across conditions led to two measures of decoder evidence per participant: one for validly predicted shapes and one for invalidly predicted shapes. This allowed us to quantify evidence for the shape as presented on the screen (by averaging evidence for validly and invalidly predicted shapes) and evidence for the cued shape [by averaging (1 − evidence) for the invalidly predicted shapes with evidence for the validly predicted shapes]. These measures were statistically tested at the group level using simple *t* tests.

For the visual cortex ROIs, voxels were selected based on the strength of the evoked response to the shapes during the shape-only runs. Other brain regions, such as the hippocampus, do not show a clear evoked response to visual stimuli. Therefore, we followed a different voxel selection procedure for the other ROIs. First, voxels were sorted by their informativeness, i.e., how different the weights for the five channels were from each other (quantified by the standard deviation of the five weights). Second, the number of voxels to include was determined by selecting between 10 and 100% of all voxels (in steps of 10%), and training and testing the model on these voxels, within the shape-only runs (i.e., training on one run and testing on the other run). For each iteration, decoding performance on shapes 2 and 4 was quantified as described above, and the number of voxels that yielded the highest decoding performance was selected (group average: hippocampus, 1536 of 3383 voxels; caudate, 590 of 2240 voxels; putamen, 1498 of 3582 voxels).

We also labeled the selected hippocampus voxels based on their subfield from the hippocampal segmentation (group average: CA1, 436 voxels; CA2–CA3–DG, 572 voxels; subiculum, 425 voxels). Differential contributions of the subfields were statistically tested by performing a one-way repeated-measures ANOVA on the measure of interest (e.g., decoding of the cued shape; Fig. 3*D*).

For the main ROI and searchlight analyses, the input to the forward model consisted of voxelwise double-gamma parameter estimates, reflecting the amplitude of the BOLD response. Additionally, decoding was also applied to the FIR model parameter estimates in visual cortex.

Searchlight analysis. To explore the specificity of presented and predicted shape representations, a multivariate searchlight approach was used to test these effects within the field of view of our functional scans (most of occipital and temporal and part of parietal and frontal cortex). A spherical searchlight with a radius of 5 voxels (7.5 mm) was passed over all functional voxels. In each searchlight, we performed shape decoding in the same manner as in the ROIs, yielding maps of decoder evidence for the presented and predicted shapes, respectively, for each participant. Group-level nonparametric permutation tests were applied to these searchlight maps using FSL Randomize (Winkler et al., 2014), correcting for multiple comparisons at $p < 0.05$ using threshold-free cluster enhancement (Smith and Nichols, 2009).

Experimental design and statistical analysis. The current study was designed to compare differences in BOLD signals evoked by validly and invalidly predicted visual stimuli. For each ROI, the amplitude and la-

tency of the BOLD response, as well as of the shape decoding signal, were quantified as described in the appropriate sections above. To test for significant differences between our conditions of interest, these measures were subjected to paired-sample *t* tests (valid vs invalid). Additionally, to test the differential involvement of the hippocampal subfields, we conducted a repeated-measures ANOVA on the predicted shape signal with the three-level factor "subfield," as described previously (see above, Shape decoding). A significant effect of subfield was followed up with planned *t* tests within individual subfields, as well as of differences between pairs of subfields. In other words, we statistically assessed subfields when this was justified by a significant omnibus effect to assess the distribution of the predicted shape signal over subfields. This hierarchical approach helps control the false positive rate, rather than simply examining all possible comparisons. Statistical testing of the whole-brain searchlight results is described in detail above (see Searchlight analysis).

Code accessibility. Data and code are available upon request from the first author (peter.kok@yale.edu).

Results

Participants were exposed to auditory tones that validly or invalidly predicted the upcoming shape stimulus (Fig. 1*A*,*B*). This first shape was followed by a second shape that was either identical to the first or slightly warped. Participants performed a shape discrimination task, reporting whether the two shapes on a given trial were the same or different.

Behavior

Participants were able to discriminate small differences in the complex shapes, during the shape-only runs (36.9 ± 2.3% modulation of the 3.18 Hz radial frequency component, mean ± SEM) and during the prediction runs (valid trials, 31.6 ± 2.5%; invalid trials, 33.2 ± 2.9% modulation). The discrimination thresholds for valid and invalid trials were not reliably different ($t_{(23)} = 1.00$, $p = 0.32$). This is not surprising, as the discrimination task was independent of the prediction manipulation: the auditory cue provided no information about which choice was correct, and the shape manipulation on different trials was orthogonal to the feature dimensions defining the shape space. Accuracy and reaction times (RTs) also did not differ significantly between valid (accuracy, 70.6 ± 1.2%; RT, 575 ± 16 ms) and invalid (accuracy, 68.8 ± 1.5%; RT, 573 ± 18 ms; both *p* values >0.20) trials, which was expected because these conditions were staircased separately to the same performance level.

Shape reconstruction

The decoder successfully reconstructed the presented shapes from the pattern of activity in visual cortex (V1: $t_{(23)} = 14.72$, $p = 3.4 \times 10^{-13}$; V2: $t_{(23)} = 14.23$, $p = 6.8 \times 10^{-13}$; LO: $t_{(23)} = 7.04$, $p = 3.5 \times 10^{-7}$), with a modest but significant modulation by the predictive cues in V1 ($t_{(23)} = 2.58$, $p = 0.017$; Fig. 3*A*) but not in V2 ($t_{(23)} = 1.42$, $p = 0.17$) or LO ($t_{(23)} = 0.17$, $p = 0.87$). In other words, shape representations in visual cortex were dominated by what was presented to the eyes.

The results were strikingly different in the hippocampus. Here, the pattern of activity contained a representation of the predicted shape ($t_{(23)} = 2.86$, $p = 0.0089$), whereas the presented shape was not significantly represented ($t_{(23)} = 0.54$, $p = 0.59$). That is, shape representations in the hippocampus were fully determined by the auditory cue and the expectation it established (Fig. 3*B*).

This dissociation was confirmed by a searchlight analysis, which revealed significant evidence for the presented shape in the occipital lobe and for the predicted (but not presented) shape in the hippocampus (Fig. 3*C*). This analysis also revealed evidence for the predicted shape in more anterior occipital cortex and a few smaller clusters elsewhere (Table 1).

Table 1. Searchlight results

Anatomical region	Hemisphere	Cluster size	Peak *p*	Coordinates (*x*, *y*, *z*)
Presented shape decoding				
Posterior occipital cortex	Bilateral	8964	<0.001	−22, −88, −22
Predicted shape decoding				
Calcarine sulcus	Right	180	0.016	14, −70, 20
Hippocampus	Right	63	0.026	24, −18, 16
Middle cingulate	Right	27	0.028	2, 8, 40
Caudate	Left	7	0.028	−16, 4, 8
Cerebellum	Left	5	0.044	−26, −62, 22

All *p* values are corrected for multiple comparisons. Coordinates reflect local maxima of significant clusters in MNI space.

Note that the hippocampal cluster in the searchlight analysis was in the right hemisphere only, whereas we collapsed over hemisphere in the ROI analysis. This apparent laterality may be an artifact of statistical thresholding or may indicate a genuine hemispheric difference. We did not have hypotheses about left versus right or anterior versus posterior hippocampus but investigated these divisions *post hoc* because of the searchlight results by subdividing the hippocampal ROI. There were no reliable differences in evidence for the predicted shape in left versus right ($p = 0.72$) or anterior versus posterior ($p = 0.93$) hippocampus. In fact, decoding of the predicted shape was significant within each of these four subdivisions of the hippocampus individually (all *p* values <0.05).

To investigate the circuitry underlying these predictions further, we applied an automated anatomical segmentation method to distinguish the subfields of the hippocampus. This analysis revealed that hippocampal subfields encoded the predicted shapes to different extents ($F_{(2,46)} = 6.45$, $p = 0.0034$; Fig. 3*D*). The CA3 subfield is thought to be most strongly involved in pattern completion (i.e., retrieving previously encoded memories from partial cues), whereas CA1 compares such retrieved memories to incoming sensory input supplied by entorhinal cortex (EC). Accordingly, we hypothesized that CA3 would have a purer representation of the predicted shape than CA1, since the latter would also be affected by the presented shape. In line with this hypothesis, predicted shapes could be reconstructed better in CA3 (combined with CA2 and dentate gyrus) than in CA1 (difference between ROIs: $t_{(23)} = 2.31$, $p = 0.031$). Note that the representation of the predicted shapes in CA2–CA3–DG itself was not statistically significant but was only a trend ($t_{(23)} = 2.04$, $p = 0.053$), whereas the effect in CA1 was far from significant and even numerically negative ($t_{(23)} = -1.08$, $p = 0.29$). Surprisingly, predicted shapes were also strongly represented in the subiculum ($t_{(23)} = 2.97$, $p = 0.0069$), which could perhaps be related to its known role in relaying hippocampal signals back to sensory cortex.

We interpret the presence of shape expectations in the hippocampus as reflecting relational memory (Cohen and Eichenbaum, 1993): item memories of the tones and shapes are bound together in a temporal relationship during the practice phase and then further during valid trials; when a particular tone cue is encountered, its item memory retrieves this relationship and reactivates the item memory for the associated shape. This framework suggests that the success of our decoder depends on the extent to which it has learned about the item memories for different shapes. We examined this hypothesis by breaking down our training examples based on familiarity with the shapes, separating the two shape-only runs rather than collapsing, as was

done in all of the analyses above. Specifically, we anticipated that training the decoder on the second shape-only run at the end of the session (run 6), after participants had the opportunity to repeatedly encode the shapes, would be more effective than training on the first shape-only run at the beginning of the session (run 1), when the shapes were more novel. Indeed, we found a significantly stronger representation of the predicted shape when the reconstruction model was trained on the last versus first shape-only run in CA2–CA3–DG ($t_{23} = 3.09, p = 0.0052$) but not in CA1 ($t_{(23)} = -0.23, p = 0.82$) or the subiculum ($t_{(23)} = 0.89, p = 0.38$). Following up on this significant difference in CA2–CA3–DG, we found that decoding of the predicted shape was significant when training the model on the last run only ($t_{(23)} = 2.99, p = 0.0065$) but not when training it on the first run ($t_{(23)} = -0.55, p = 0.59$). In visual cortex, on the other hand, training on the last versus first shape-only run did not affect the representation of the predicted shape (V1: $t_{(23)} = -0.88$, $p = 0.39$; V2: $t_{(23)} = 0.17$, $p = 0.87$; LO: $t_{(23)} = 0.004$, $p = 0.997$).

Visual facilitation

As reported above, shape representations in visual cortex were dominated by the shapes presented to the eyes. However, the temporal evolution of these representations was strongly affected by the auditory prediction cues. We characterized the time courses of both the mean BOLD response and the shape decoding signal by fitting a canonical (double-gamma) hemodynamic response function (HRF) and its temporal derivative. The parameter estimate of the canonical HRF indicates the peak amplitude of the signal, whereas the temporal derivative parameter estimate reflects the latency of the signal (Friston et al., 1998; Henson et al., 2002).

This approach revealed that there was a modest but highly reliable difference in the latency of the BOLD response evoked by validly and invalidly predicted shapes, as measured by the temporal derivative (V1: $t_{(23)} = 6.33, p = 1.9 \times 10^{-6}$; V2: $t_{(23)} = 7.31$, $p = 1.9 \times 10^{-7}$; LO: $t_{(23)} = 7.48$, $p = 1.32 \times 10^{-7}$; Fig. 4*A*). In other words, the BOLD response in visual cortex was significantly delayed by invalid auditory cues. Note that this was not caused by the stimuli per se, as the tone–shape mappings were arbitrary and reversed halfway through the study. There was no significant difference in the amplitude of the BOLD response between conditions, as measured by the canonical HRF (V1: $t_{(23)} = 0.84$, $p = 0.41$; V2: $t_{(23)} = 1.64$, $p = 0.12$; LO: $t_{(23)} = 1.96$, $p = 0.063$).

The delay for invalidly predicted shapes was also apparent in the temporal evolution of the reconstructed shape representations (Fig. 4*B*). There was a reliable difference in the temporal derivative of the time-resolved decoding signal in V1 ($t_{(23)} = 3.40$, $p = 0.0024$) and V2 ($t_{(23)} = 3.06$, $p = 0.0056$), with a marginal effect in LO ($t_{(23)} = 1.96, p = 0.062$). In V1, the peak of the decoding signal was significantly lower for invalidly predicted shapes than for validly predicted shapes ($t_{(23)} = 2.73, p = 0.012$), whereas there was no such effect in V2 ($t_{(23)} = 1.11, p = 0.27$) or LO ($t_{(23)} = 0.15, p = 0.87$).

In summary, although there was a modest effect of prediction on the amplitude of the shape decoding signal in V1, the most striking effects of prediction in visual cortex were on the latency of the BOLD response and decoding signal.

Hippocampal–cortical relationships

It is impossible with fMRI to establish that hippocampal prediction causes visual facilitation, but a precondition for such a mechanism is that these two measures should be related. Testing this relationship within participants was not possible in the current study because single-trial reconstruction and decoding of predictions was too noisy, especially in the hippocampus. We thus adopted an across-participant approach: we hypothesized that participants with greater decoding of the predicted shape in the hippocampus should have a greater latency shift in the decoding of invalidly versus validly cued shapes in visual cortex. We found such a relationship between the hippocampus and LO ($r = 0.42$, $p = 0.040$; Fig. 5), but not with V1 ($r = -0.29$, $p = 0.17$) or V2 ($r = -0.05$, $p = 0.81$).

Strikingly, the hippocampal–LO relationship differed strongly across hippocampal subfields (Fig. 5). In CA2–CA3–DG, as for the hippocampus as a whole, there was a reliable positive relationship ($r = 0.59, p = 0.0022$), with more hippocampal prediction associated with more LO facilitation. In CA1, however, the relationship was reliably negative ($r = -0.44, p = 0.032$), with more hippocampal prediction associated with less LO facilitation. There was no reliable relationship in subiculum ($r = 0.09, p = 0.66$). The surprising negative relationship between CA1 and LO also held for V1 ($r = -0.55, p = 0.0047$) and V2 ($r = -0.57, p = 0.0036$), whereas the positive relationship of CA2–CA3–DG was found only for LO (V1: $r = 0.039, p = 0.86$; V2: $r = 0.17, p = 0.43$).

In summary, although these findings do not resolve the causal direction of hippocampal–cortical interactions, they are consistent with the proposed mechanism of the hippocampus supplying predictions to visual cortex, at least more than if we had found no such relationships.

Caudate predictions

In addition to the hippocampus, we also examined the striatum, specifically the caudate and putamen, based on previous studies of prediction (den Ouden et al., 2009; Turk-Browne et al., 2009) as well as the known involvement of the striatum in associative learning (Poldrack et al., 2001; Shohamy and Turk-Browne, 2013). We found that the caudate represented the predicted shape ($t_{(23)} = 3.07$, $p = 0.0054$) but not the presented shape ($t_{(23)} = 0.10$, $p = 0.92$), as in the hippocampus. We could not reconstruct shape information from the putamen for either the predicted ($t_{(23)} = 1.58$, $p = 0.13$) or the presented ($t_{(23)} = 0.27$, $p = 0.79$) shapes. Unlike the hippocampus, LO facilitation did not correlate with prediction in the caudate ($r = 0.23, p = 0.27$) or putamen ($r = 0.23, p = 0.27$), nor were there correlations with V1 or V2 (p values >0.05).

Learning and contingency reversal

All expectations in this study were learned during the session, raising the interesting question of how they evolve over the learning process. Unfortunately, the current design was not well suited to answer this question because participants were pretrained on both contingencies during practice runs without fMRI. Because of these long practice runs, we did not anticipate much additional learning to take place during fMRI acquisition. Nevertheless, we examined whether the strength of prediction changed over time by repeating the main analysis separately within each of the two blocks (first and second halves, respectively) of both contingencies. We analyzed potential effects of "block" (first vs second half of each run) and "contingency" (before vs after contingency reversal) using a two-way repeated-measures ANOVA. Collapsing across contingencies, we did not find evidence for an increase in prediction signals from the first to the second block in either the hippocampus (main effect of block: $F_{(1,22)} = 0.005, p = 0.94$) or the caudate ($F_{(1,22)} = 0.26, p = 0.61$). There was also no significant difference in prediction signals before versus after the contingency reversal, collapsing over blocks, in hippocampus

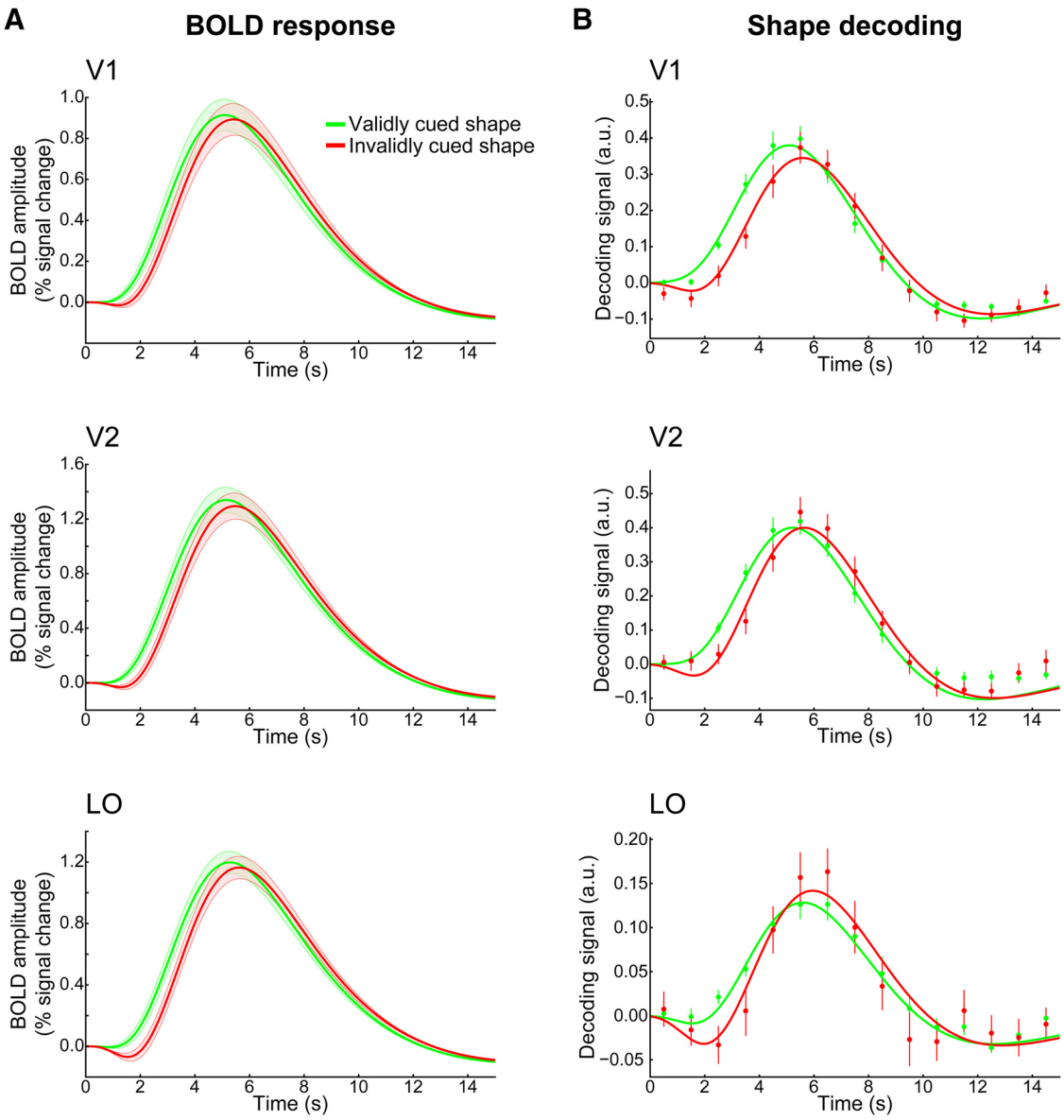

Figure 4. Time-resolved activity and decoding in visual cortex. ***A***, Time course of the mean BOLD response, separately for validly (green) and invalidly (red) predicted shapes, in visual cortex. These time courses reflect the fit of the canonical HRF and its temporal derivative to the preprocessed fMRI data by condition. ***B***, Time course of the shape decoding signal, separately for validly (green) and invalidly (red) predicted shapes. Here, the canonical HRF and its derivative were not fit to the fMRI data directly, but rather to a continuous decoding signal obtained by reconstructing shape information for each time point from FIR parameter estimates. Shaded regions and error bars indicate SEM.

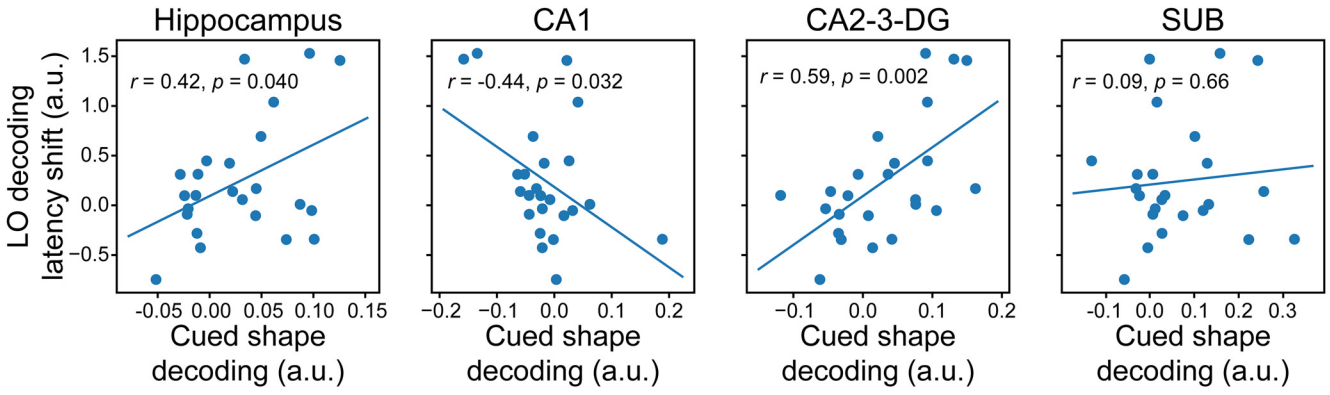

Figure 5. Hippocampal–cortical interactions. Correlation between the strength of predicted shape decoding in the hippocampus and the latency shift in the decoding signal caused by predictions in LO.

(main effect of contingency: $F_{(1,22)} = 0.62$, $p = 0.44$) and caudate ($F_{(1,22)} = 3.83$, $p = 0.063$), although note that caudate showed a nonsignificant trend toward prediction signals being stronger after the contingency reversal than before. Finally, there was no interaction between block and contingency in either region (both p values >0.20). In short, we did not observe gradual learning of the prediction signals (beyond the practice runs) in either hippocampus or caudate.

Same versus different trials

In 50% of the trials the two shapes were identical, and in the other 50% the second shape was slightly warped with respect to the first. For our main analyses, we collapsed across these two trial types. In an additional analysis, we investigated whether or not expectation effects differed for same versus different trials. We found that, in line with the warps being very subtle and orthogonal to our shape space (see Materials and Methods), shape decoding in visual cortex did not differ between same and different trials (V1: $F_{(1,22)} = 0.33$, $p = 0.57$; V2: $F_{(1,22)} = 0.17$, $p = 0.69$; LO: $F_{(1,22)} = 0.73$, $p = 0.40$), nor did the effect of same versus different interact with the validity of the expectation cue (V1: $F_{(1,22)} = 0.14$, $p = 0.71$; V2: $F_{(1,22)} < 0.01$, $p = 0.98$; LO: $F_{(1,22)} = 0.73$, $p = 0.40$). In hippocampus, decoding of the predicted shape was not affected by the same versus different distinction ($F_{(1,22)} = 0.25$, $p = 0.62$). However, there was a trend toward decoding of the shape presented on screen being slightly better for same than for different trials ($F_{(1,22)} = 4.22$, $p = 0.052$). In caudate, the reverse applied; decoding of the presented shape was not affected by the same versus different distinction ($F_{(1,22)} = 0.02$, $p = 0.89$), yet decoding of the predicted shape was slightly better for same than different trials ($F_{(1,22)} = 4.59$, $p = 0.043$). In summary, unlike the caudate, representations of the predicted shapes in the hippocampus were immune to variation in visual input, despite also registering this information, as reflected in representations of the presented shape.

Discussion

Predictive coding theories (Mumford, 1992; Rao and Ballard, 1999; Friston, 2005) of cortical processing fit well with the strong influence of predictions on sensory processing reported here and elsewhere (for review, see den Ouden et al., 2012; Summerfield and de Lange, 2014). Generally, such models seek to explain mostly lower-level phenomena that can be resolved within local circuits of visual cortex, such as end-stopping and surround suppression (Rao and Ballard, 1999; Spratling, 2010). However, many predictive cues in our environment require cross-modal interactions and invoke complex expectations about objects. We hypothesized that such cross-modal predictions may be generated in the hippocampus. Specifically, after presentation of a predictive cue, CA3 may retrieve the associated item through pattern completion of a learned temporal relationship and send this prediction to CA1 and from there back to sensory cortex, including through the subiculum (Lavenex and Amaral, 2000; Roy et al., 2017). Within CA1, memory-based predictions (originating from CA3) have been proposed to inhibit matching sensory signals (from EC), thereby signaling novelty or prediction error (Lisman and Grace, 2005; Kumaran and Maguire, 2007; Chen et al., 2011, 2015; Duncan et al., 2012). Based on this model, one would expect CA3, but not CA1, to represent only the predicted item, and that is indeed what we observed in the current study. In addition, a representation of the predicted item was found in the subiculum, known to be a major relay between the hippocampus and sensory cortex, though admittedly not well understood and often excluded from hippocampal models (McClelland et al., 1995; Schapiro et al., 2017).

If the hippocampus is a source of sensory expectations, there should be a relationship between the strength of hippocampal predictions and effects of prediction in visual cortex. Although the current study did not allow us to study this relationship within participant (across trials) because of poor single-trial decoding, we did find such a relationship across participants. This also held for the CA2–CA3–DG subfield alone, in line with CA3's proposed role in generating predictions via pattern completion. Conversely, in line with the proposed inhibitory role of predictions in CA1, prediction strength in this subfield was negatively correlated with the facilitative effects of prediction on visual cortex.

The model outlined above suggests a specific direction of neural signal flow during the generation of predictions, namely from CA3 through CA1 and the subiculum to cortex. However, because of the slow nature of the hemodynamic response and the lack of causal intervention, standard fMRI does not allow us to distinguish the direction of flow between regions. Future studies will be needed to directly address this important issue, including with intracranial recordings in neurological patients with both depth electrodes in the hippocampus and surface electrodes in sensory cortex. Additionally, signals from the hippocampus to cortex are known to arrive in the deep layers of EC, whereas signals from cortex to hippocampus flow through the superficial layers of EC (Lavenex and Amaral, 2000). Using high-field fMRI to study layer-specific prediction signals in EC could thus be used to help establish the direction of signal flow between hippocampus and cortex (Maass et al., 2014; Muckli et al., 2015; Kok et al., 2016a).

The current study suggests that the hippocampus is involved in signaling cross-modal predictions. However, there are several other mechanisms for prediction in the brain, including related to object recognition and semantic labels in medial prefrontal cortex (Bar et al., 2006) and to value and reinforcement learning in the ventral striatum (den Ouden et al., 2012), as well as many other areas of polymodal association cortex that receive the required sensory inputs. What might distinguish the contribution of the hippocampus is the ability to quickly and flexibly learn new predictions, whereas these other systems learn more gradually after extensive experience and consolidation (McClelland et al., 1995; Schapiro et al., 2017). Regardless, further work will be needed to understand the relative contributions of each system and whether they have a cooperative or competitive relationship. For instance, it has been proposed that the striatum serves as a gating mechanism, upregulating connectivity between top-down attention systems and sensory cortex when prediction errors occur (Zink et al., 2006; den Ouden et al., 2010), rather than containing actual stimulus representations itself. However, our findings suggest that at least the caudate contains shape-specific representations of predicted stimuli and that they are encoded in similar activity patterns to the corresponding sensory stimuli (given generalization of the model from shape-only to prediction runs). An important avenue for future research would be to tease apart the roles of these two learning systems, the hippocampus and the striatum, in storing predictive associations and their respective roles in sending feedback to sensory cortex (Poldrack et al., 2001; Shohamy and Turk-Browne, 2013). The current study only investigated the consequences of learning such associations, rather than the learning process itself, since fMRI acquisition was preceded by a practice phase that familiarized participants with the associations many times. In future work, fMRI signals from

the hippocampus and striatum could be acquired while the associations are being learned, to establish in which of these regions the time course of learning best matches the build-up of expectation signals in visual cortex.

The effects of the complex shape predictions on processing in the visual cortex, as reported here, differ strikingly from those reported previously for low-level feature predictions, using an otherwise similar paradigm (Kok et al., 2012). Whereas invalid grating orientation predictions in that study led to both an increased peak BOLD amplitude and a reduced orientation representation in V1 (Kok et al., 2012), the current study found that invalid shape predictions lead to delayed signals, both in terms of BOLD amplitude and shape representations. Although the cause of this difference is currently unclear, we offer a couple of potential explanations. First, predictions about low-level features and complex shapes may be encoded differently in visual cortex. Whereas a prediction about grating orientation could be encoded by simply increasing the gain of all neurons tuned for that orientation across the visual field (Kok et al., 2016b), complex shape predictions would require encoding different orientations and curvatures at specific retinotopic locations. This may be a particularly difficult challenge given that complex shapes are known to be encoded in a spatially invariant manner in higher-level visual cortex (DiCarlo et al., 2012). In line with this account, there is evidence that predictions about low-level features and complex natural images can have different effects on perception (Denison et al., 2011, 2016). Second, whereas invalid gratings in the study by Kok et al. (2012) were maximally different from predicted gratings (i.e., orthogonal orientations), the difference between predicted and unpredicted shapes was more subtle. Such small violations may be less prone to strong prediction errors but may rather lead to an integration of top-down predictions and bottom-up sensory signals (Kok et al., 2013). Clearly, future research is required to investigate these and other factors. A clear next step is to investigate how low-level feature predictions, particularly when involving cross-modal cues, engage the hippocampus.

Potentially, the latency differences between visual cortex signals induced by validly and invalidly predicted shapes might be the result of a suppression of neural signals evoked by the first shape on a given trial by an invalid prediction, but less so for the second shape (which is no longer really unexpected once the first shape has been observed). This could lead to a delayed peak activity once convolved with the BOLD response. This scenario seems particularly plausible for the reconstructed shape representations, since the early BOLD signals on invalid trials would presumably contain a mixture of the predicted and presented shapes, which might, to some extent, cancel each other out in the eyes of the decoder.

Previous studies have found that predictive cues can lead to the cortical reinstatement of expected stimuli (Kok et al., 2014; Hindy et al., 2016), in anticipation of the actual sensory inputs (Kok et al., 2017). Such cortical reinstatement has been shown for other cognitive processes as well, such as visual short-term memory (Harrison and Tong, 2009), mental imagery (Stokes et al., 2009; Albers et al., 2013), and preparatory attention (Peelen and Kastner, 2011; Myers et al., 2015). One intriguing possibility is that these different cognitive processes are subserved by the same neural mechanism (Pearson and Westbrook, 2015). Specifically, is cortical reinstatement in working memory, imagery, and attention mediated by the hippocampus, as it seems to be in associative memory (Bosch et al., 2014; Gordon et al., 2014) and the cross-modal predictions studied here? Additionally, do these different processes affect visual cortex the same way, or do different processes modulate different layers of visual cortex, in support of different computational goals (Friston, 2005; Muckli et al., 2015; Kok et al., 2016a)?

In conclusion, here we find that patterns of neural activity in the hippocampus reflect stimulus-specific predictions, as signaled by cross-modal cues. Furthermore, the strength of these hippocampal signals correlates with facilitation of perceptual processing in visual cortex. These findings help bridge the gap between memory and sensory systems in the human brain.

References

Albers AM, Kok P, Toni I, Dijkerman HC, de Lange FP (2013) Shared representations for working memory and mental imagery in early visual cortex. Curr Biol 23:1427–1431. CrossRef Medline

Alink A, Schwiedrzik CM, Kohler A, Singer W, Muckli L (2010) Stimulus predictability reduces responses in primary visual cortex. J Neurosci 30:2960–2966. CrossRef Medline

Aly M, Turk-Browne NB (2016a) Attention promotes episodic encoding by stabilizing hippocampal representations. Proc Natl Acad Sci U S A 113:E420–E429. CrossRef Medline

Aly M, Turk-Browne NB (2016b) Attention stabilizes representations in the human hippocampus. Cereb Cortex 26:783–796. Medline

Andersson JL, Skare S, Ashburner J (2003) How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage 20:870–888. CrossRef Medline

Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmidt AM, Dale AM, Hämäläinen MS, Marinkovic K, Schacter DL, Rosen BR, Halgren E (2006) Top-down facilitation of visual recognition. Proc Natl Acad Sci U S A 103:449–454. CrossRef Medline

Bell AH, Summerfield C, Morin EL, Malecek NJ, Ungerleider LG (2016) Encoding of stimulus probability in macaque inferior temporal cortex. Curr Biol 26:2280–2290. CrossRef Medline

Bosch SE, Jehee JF, Fernandez G, Doeller CF (2014) Reinstatement of associative memories in early visual cortex is signaled by the hippocampus. J Neurosci 34:7493–7500. CrossRef Medline

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

Brouwer GJ, Heeger DJ (2009) Decoding and reconstructing color from responses in human visual cortex. J Neurosci 29:13992–14003. CrossRef Medline

Brouwer GJ, Heeger DJ (2011) Cross-orientation suppression in human visual cortex. J Neurophysiol 106:2108–2119. CrossRef Medline

Carr VA, Rissman J, Wagner AD (2010) Imaging the human medial temporal lobe with high-resolution fMRI. Neuron 65:298–308. CrossRef Medline

Chen J, Olsen RK, Preston AR, Glover GH, Wagner AD (2011) Associative retrieval processes in the human medial temporal lobe: hippocampal retrieval success and CA1 mismatch detection. Learn Mem 18:523–528. CrossRef Medline

Chen J, Cook PA, Wagner AD (2015) Prediction strength modulates responses in human area CA1 to sequence violations. J Neurophysiol 114:1227–1238. CrossRef Medline

Chun MM, Phelps EA (1999) Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. Nat Neurosci 2:844–847. CrossRef Medline

Cohen NJ, Eichenbaum H (1993) Memory, amnesia, and the hippocampal system. Cambridge, MA: MIT.

Davachi L (2006) Item, context and relational episodic encoding in humans. Curr Opin Neurobiol 16:693–700. CrossRef Medline

Davachi L, DuBrow S (2015) How the hippocampus preserves order: the role of prediction and context. Trends Cogn Sci 19:92–99. CrossRef Medline

den Ouden HE, Friston KJ, Daw ND, McIntosh AR, Stephan KE (2009) A dual role for prediction error in associative learning. Cereb Cortex 19:1175–1185. CrossRef Medline

den Ouden HE, Daunizeau J, Roiser J, Friston KJ, Stephan KE (2010) Striatal prediction error modulates cortical coupling. J Neurosci 30:3210–3219. CrossRef Medline

den Ouden HE, Kok P, De Lange FP (2012) How prediction errors shape perception, attention, and motivation. Front Psychol 3:548. Medline

Denison RN, Piazza EA, Silver MA (2011) Predictive context influences perceptual selection during binocular rivalry. Front Hum Neurosci 5:166. Medline

Denison RN, Sheynin J, Silver MA (2016) Perceptual suppression of predicted natural images. J Vis 16:6. CrossRef Medline

DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? Neuron 73:415–434. CrossRef Medline

Drucker DM, Aguirre GK (2009) Different spatial scales of shape similarity representation in lateral and ventral LOC. Cereb Cortex 19:2269–2280. CrossRef Medline

Duncan K, Ketz N, Inati SJ, Davachi L (2012) Evidence for area CA1 as a match/mismatch detector: a high-resolution fMRI study of the human hippocampus. Hippocampus 22:389–398. CrossRef Medline

Duvernoy HM (2005) The human hippocampus: functional anatomy, vascularization and serial sections with MRI. Berlin, Germany Springer Science and Business Media.

Eichenbaum H, Fortin NJ (2009) The neurobiology of memory based predictions. Philos Trans R Soc Lond B Biol Sci 364:1183–1191. CrossRef Medline

Eichenbaum H, Yonelinas AP, Ranganath C (2007) The medial temporal lobe and recognition memory. Annu Rev Neurosci 30:123–152. CrossRef Medline

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cereb cortex. Cereb Cortex 1:1–47. CrossRef Medline

Friston K (2005) A theory of cortical responses. Philos Trans R Soc B Biol Sci 360:815–836. CrossRef

Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R (1998) Event-related fMRI: characterizing differential responses. Neuroimage 7:30–40. CrossRef Medline

Garvert MM, Dolan RJ, Behrens TE (2017) A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. eLife 6:e17086. CrossRef Medline

Gordon AM, Rissman J, Kiani R, Wagner AD (2014) Cortical reinstatement mediates the relationship between content-specific encoding activity and subsequent recollection decisions. Cereb Cortex 24:3350–3364. CrossRef Medline

Hannula DE, Tranel D, Cohen NJ (2006) The long and the short of it: relational memory impairments in amnesia, even at short lags. J Neurosci 26:8352–8359. CrossRef Medline

Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. Nature 458:632–635. CrossRef Medline

Henke K (2010) A model for memory systems based on processing modes rather than consciousness. Nat Rev Neurosci 11:523–532. CrossRef Medline

Henson RN, Price CJ, Rugg MD, Turner R, Friston KJ (2002) Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. Neuroimage 15:83–97. CrossRef Medline

Hindy NC, Ng FY, Turk-Browne NB (2016) Linking pattern completion in the hippocampus to predictive coding in visual cortex. Nat Neurosci 19:665–667. CrossRef Medline

Hsieh LT, Gruber MJ, Jenkins LJ, Ranganath C (2014) Hippocampal activity patterns carry information about objects in temporal context. Neuron 81:1165–1178. CrossRef Medline

Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM (2012) FSL. Neuroimage 62:782–790. CrossRef Medline

Kaposvari P, Kumar S, Vogels R (2016) Statistical learning signals in macaque inferior temporal cortex. Cereb Cortex 28:250–266. CrossRef Medline

Kok P, de Lange FP (2014) Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. Curr Biol 24:1531–1535. CrossRef Medline

Kok P, Jehee JF, De Lange FP (2012) Less is more: expectation sharpens representations in the primary visual cortex. Neuron 75:265–270. CrossRef Medline

Kok P, Brouwer GJ, Van Gerven MA, de Lange FP (2013) Prior expectations bias sensory representations in visual cortex. J Neurosci 33:16275–16284. CrossRef Medline

Kok P, Failing MF, De Lange FP (2014) Prior expectations evoke stimulus templates in the primary visual cortex. J Cogn Neurosci 26:1546–1554. CrossRef Medline

Kok P, Bains LJ, van Mourik T, Norris DG, de Lange FP (2016a) Selective activation of the deep layers of the human primary visual cortex by top-down feedback. Curr Biol 26:371–376. CrossRef Medline

Kok P, van Lieshout LL, de Lange FP (2016b) Local expectation violations result in global activity gain in primary visual cortex. Sci Rep 6:37706. CrossRef Medline

Kok P, Mostert P, de Lange FP (2017) Prior expectations induce prestimulus sensory templates. Proc Natl Acad Sci U S A 114:10473–10478. CrossRef Medline

Konkel A, Warren DE, Duff MC, Tranel DN, Cohen NJ (2008) Hippocampal amnesia impairs all manner of relational memory. Front Hum Neurosci 2:15. CrossRef Medline

Kumaran D, Maguire EA (2007) Which computational mechanisms operate in the hippocampus during novelty detection? Hippocampus 17:735–748. CrossRef Medline

Lavenex P, Amaral DG (2000) Hippocampal-neocortical interaction: a hierarchy of associativity. Hippocampus 10:420–430. CrossRef Medline

Lee TS, Nguyen M (2001) Dynamics of subjective contour formation in the early visual cortex. Proc Natl Acad Sci U S A 98:1907–1911. CrossRef Medline

Lisman JE, Grace AA (2005) The hippocampal-VTA loop: controlling the entry of information into long-term memory. Neuron 46:703–713. CrossRef Medline

Maass A, Schütze H, Speck O, Yonelinas A, Tempelmann C, Heinze H-J, Berron D, Cardenas-Blanco A, Brodersen KH, Stephan KE, Düzel E (2014) Laminar activity in the hippocampus and entorhinal cortex related to novelty and episodic encoding. Nat Commun 5:5547. CrossRef Medline

McClelland JL, McNaughton BL, O'Reilly RC (1995) Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol Rev 102:419–457. CrossRef Medline

Meyer T, Olson CR (2011) Statistical learning of visual transitions in monkey inferotemporal cortex. Proc Natl Acad Sci U S A 108:19401–19406. CrossRef Medline

Muckli L, De Martino F, Vizioli L, Petro LS, Smith FW, Ugurbil K, Goebel R, Yacoub E (2015) Contextual feedback to superficial layers of V1. Curr Biol 25:2690–2695. CrossRef Medline

Mumford D (1992) On the computational architecture of the neocortex. Biol Cybern 66:241–251. CrossRef Medline

Myers NE, Rohenkohl G, Wyart V, Woolrich MW, Nobre AC, Stokes MG (2015) Testing sensory evidence against mnemonic templates. eLife 4:e09000. Medline

Op de Beeck H, Wagemans J, Vogels R (2001) Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. Nat Neurosci 4:1244–1252. CrossRef Medline

Pearson J, Westbrook F (2015) Phantom perception: voluntary and involuntary nonretinal vision. Trends Cogn Sci 19:278–284. CrossRef Medline

Peelen MV, Kastner S (2011) A neural basis for real-world visual search in human occipitotemporal cortex. Proc Natl Acad Sci U S A 108:12125–12130. CrossRef Medline

Poldrack RA, Clark J, Paré-Blagoev EJ, Shohamy D, Creso Moyano J, Myers C, Gluck MA (2001) Interactive memory systems in the human brain. Nature 414:546–550. CrossRef Medline

Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87. CrossRef Medline

Roy DS, Kitamura T, Okuyama T, Ogawa SK, Sun C, Obata Y, Yoshiki A, Tonegawa S (2017) Distinct neural circuits for the formation and retrieval of episodic memories. Cell 170:1000–1012.e19. CrossRef Medline

Schapiro AC, Kustner LV, Turk-Browne NB (2012) Shaping of object representations in the human medial temporal lobe based on temporal regularities. Curr Biol 22:1622–1627. CrossRef Medline

Schapiro AC, Gregory E, Landau B, McCloskey M, Turk-Browne NB (2014) The necessity of the medial temporal lobe for statistical learning. J Cogn Neurosci 26:1736–1747. CrossRef Medline

Schapiro AC, Turk-Browne NB, Botvinick MM, Norman KA (2017) Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. Philos Trans R Soc Lond B Biol Sci 372:20160049. CrossRef Medline

Shohamy D, Turk-Browne NB (2013) Mechanisms for widespread hippocampal involvement in cognition. J Exp Psychol Gen 142:1159–1170. CrossRef Medline

Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44:83–98. CrossRef Medline

Solomon PR, Vander Schaaf ER, Thompson RF, Weisz DJ (1986) Hippocampus and trace conditioning of the rabbit's classically conditioned nictitating membrane response. Behav Neurosci 100:729–744. CrossRef Medline

Spratling MW (2010) Predictive coding as a model of response properties in cortical area V1. J Neurosci 30:3531–3543. CrossRef Medline

Staresina BP, Davachi L (2009) Mind the gap: binding experiences across space and time in the human hippocampus. Neuron 63:267–276. CrossRef Medline

Stokes M, Thompson R, Cusack R, Duncan J (2009) Top-down activation of shape-specific population codes in visual cortex during mental imagery. J Neurosci 29:1565–1572. CrossRef Medline

Summerfield C, de Lange FP (2014) Expectation in perceptual decision making: neural and computational mechanisms. Nat Rev Neurosci 15:745–756. CrossRef Medline

Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T (2008) Neural repetition suppression reflects fulfilled perceptual expectations. Nat Neurosci 11:1004–1006. CrossRef Medline

Sutherland RJ, McDonald RJ, Hill CR, Rudy JW (1989) Damage to the hippocampal formation in rats selectively impairs the ability to learn cue relationships. Behav Neural Biol 52:331–356. CrossRef Medline

Todorovic A, van Ede F, Maris E, de Lange FP (2011) Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. J Neurosci 31:9118–9123. CrossRef Medline

Treves A, Rolls ET (1994) Computational analysis of the role of the hippocampus in memory. Hippocampus 4:374–391. CrossRef Medline

Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK (2009) Neural evidence of statistical learning: efficient detection of visual regularities without awareness. J Cogn Neurosci 21:1934–1945. CrossRef Medline

Wacongne C, Labyt E, van Wassenhove V, Bekinschtein T, Naccache L, Dehaene S (2011) Evidence for a hierarchy of predictions and prediction errors in human cortex. Proc Natl Acad Sci U S A 108:20754–20759. CrossRef Medline

Wallenstein GV, Eichenbaum H, Hasselmo ME (1998) The hippocampus as an associator of discontiguous events. Trends Neurosci 21:317–323. CrossRef Medline

Watson AB, Pelli DG (1983) QUEST: a Bayesian adaptive psychometric method. Percept Psychophys 33:113–120. CrossRef Medline

Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE (2014) Permutation inference for the general linear model. Neuroimage 92:381–397. CrossRef Medline

Yushkevich PA, Pluta JB, Wang H, Xie L, Ding SL, Gertje EC, Mancuso L, Kliot D, Das SR, Wolk DA (2015) Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment: automatic morphometry of MTL subfields in MCI. Hum Brain Mapp 36:258–287. CrossRef Medline

Zahn CT, Roskies RZ (1972) Fourier descriptors for plane closed curves. IEEE Trans Comput C 21:269–281.

Zink CF, Pagnoni G, Chappelow J, Martin-Skurski M, Berns GS (2006) Human striatal activation reflects degree of stimulus saliency. Neuroimage 29:977–983. CrossRef Medline