Content-based Dissociation of Hippocampal Involvement in Prediction

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

Abstract

■ Recent work suggests that a key function of the hippocampus is to predict the future. This is thought to depend on its ability to bind inputs over time and space and to retrieve upcoming or missing inputs based on partial cues. In line with this, previous research has revealed prediction-related signals in the hippocampus for complex visual objects, such as fractals and abstract shapes. Implicit in such accounts is that these computations in the hippocampus reflect domain-general processes that apply across different types and modalities of stimuli. An alternative is that the hippocampus plays a more domain-specific role in predictive processing, with the type of stimuli being predicted determining its involvement. To investigate this, we compared hippocampal responses to auditory cues predicting abstract shapes (Experiment 1) versus oriented gratings (Experiment 2). We measured brain activity in male and female human participants using high-resolution fMRI, in combination with inverted encoding models to reconstruct shape and orientation information. Our results revealed that expectations about shape and orientation evoked distinct representations in the hippocampus. For complex shapes, the hippocampus represented which shape was expected, potentially serving as a source of top–down predictions. In contrast, for simple gratings, the hippocampus represented only unexpected orientations, more reminiscent of a prediction error. We discuss several potential explanations for this content-based dissociation in hippocampal function, concluding that the computational role of the hippocampus in predictive processing may depend on the nature and complexity of stimuli. ■

INTRODUCTION

Sensory processing is strongly influenced by prior expectations (De Lange, Heilbron, & Kok, 2018). Expectations about both simple features (e.g., orientation; Jabar, Filipowicz, & Anderson, 2017; Kok, Mostert, & De Lange, 2017; Kok, Jehee, & De Lange, 2012) and complex objects (e.g., shapes; Kaposvari, Kumar, & Vogels, 2018; Manahova, Mostert, Kok, Schoffelen, & De Lange, 2018; Richter, Ekman, & De Lange, 2018; Utzerath, St John-Saaltink, Buitelaar, & De Lange, 2017; Meyer & Olson, 2011) modulate processing in visual cortex. However, it is unclear whether these two kinds of expectations arise from the same top–down sources and operate via the same underlying mechanisms.

Previous research has revealed prediction-related signals in the hippocampus for complex visual objects, such as fractals (Hindy, Ng, & Turk-Browne, 2016; Schapiro, Kustner, & Turk-Browne, 2012) and abstract shapes (Kok & Turk-Browne, 2018; Wang, Shen, Tino, Welchman, & Kourtzi, 2017). These studies used fMRI to reveal that the pattern of activity in the hippocampus contains information about expected visual objects upon presentation of a predictive cue. Based on this, it has been suggested that the hippocampus may generate perceptual expectations, especially when these predictions result from rapidly learned associations between arbitrary stimuli (Schapiro, Turk-Browne, Botvinick, & Norman, 2017; Hindy et al., 2016; Davachi & DuBrow, 2015; McClelland, McNaughton, & O'Reilly, 1995). The role of the hippocampus may be particularly relevant when the associations are cross-modal, given its bidirectional connectivity with all sensory systems (Lavenex & Amaral, 2000).

This perspective raises the possibility that the hippocampus implements general-purpose computations that subserve all kinds of (associative) prediction (Buzsáki & Tingley, 2018; Stachenfeld, Botvinick, & Gershman, 2017; Lisman & Redish, 2009). That is, upon presentation of a predictive cue or context, the hippocampus may retrieve the associated outcome through pattern completion (Henke, 2010; McClelland et al., 1995), regardless of the exact nature of the stimuli. This is in line with evidence that the hippocampus is involved in many different types of predictions, pertaining to, for example, faces and scenes (Turk-Browne, Scholl, Johnson, & Chun, 2010), auditory sequences (Recasens, Gross, & Uhlhaas, 2018), odors (Eichenbaum & Fortin, 2009), and spatial locations (Liu, Sibille, & Dragoi, 2018; Stachenfeld et al., 2017).

However, an alternative and untested hypothesis is that the hippocampus only generates predictions of complex stimuli. Here, we define complex stimuli as conjunctions of features (such as oriented lines) that cannot be reduced to the sum of their parts. For instance, a triangle consists of

[1]Yale University, [2]University College London

the conjunction of three oriented lines intersecting at specific locations; if the lines were arranged differently, the triangle would cease to be. This definition of complexity is mirrored by the hierarchical organization of the visual processing pathway, with neural responses being tuned to simple visual features in early stages, and to increasingly complex (i.e., conjunctive) objects in later cortical stages (Cowell, Leger, & Serences, 2017). Visual processing in areas of the medial-temporal lobe (MTL) most directly connected to the hippocampus, such as perirhinal and parahippocampal cortices, is dominated by high-level objects and scenes, respectively (Martin, Douglas, Newsome, Man, & Barense, 2018; Murray, Bussey, & Saksida, 2007; Epstein & Kanwisher, 1998), as well as their spatial, temporal, and associative relations (Tsao et al., 2018; Garvert, Dolan, & Behrens, 2017; Hafting, Fyhn, Molden, Moser, & Moser, 2005). Processing in these regions is thought to be abstracted away from low-level sensory features (Murray et al., 2007; Lavenex & Amaral, 2000), such as orientation and pitch. It has been suggested that processing in the hippocampus is defined by conjunctive coding of these MTL representations, explaining its role in representing complex stimuli such as events, sequences, and spatial maps (Cowell, Barense, & Sadil, 2019; Behrens et al., 2018; Yonelinas, 2013). Theories casting sensory processing as hierarchical Bayesian inference (Friston, 2005; Lee & Mumford, 2003; Rao & Ballard, 1999) suggest that each brain region provides predictions only to those lower order region(s) with which it has direct feedback connections, rather than bridging the full hierarchy. Therefore, given the high-level selectivity of MTL cortex, hippocampal predictions may only traffic in complex visual stimuli and not in low-level sensory features.

In a recent study, we revealed hippocampal representations of visual predictions by exposing human participants to complex auditory cues predicting the shape of an abstract Fourier descriptor (Experiment 1, $n = 24$; Kok & Turk-Browne, 2018). Here, we tested whether hippocampal involvement is dependent on the nature of the predicted stimulus by replacing the complex shapes of Experiment 1 with simple oriented gratings (Experiment 2, $n = 24$). In both studies, we measured brain activity using high-resolution fMRI and used inverted encoding models (Brouwer & Heeger, 2009) to reconstruct shape and orientation information. Keeping the experimental design and analysis methods nearly identical between experiments allowed us to directly compare the neural effects of complex shape and low-level orientation predictions. To preview, we found that predictions about orientation and shape were represented qualitatively differently in the hippocampus.

METHODS

Participants

For both experiments, we planned to obtain a sample of 24 participants. The effect size of interest was not known in advance, so this sample size was chosen to match previous fMRI studies in our lab investigating effects of predictions in hippocampus and MTL (Hindy & Turk-Browne, 2016; Hindy et al., 2016).

Experiment 1 enrolled 25 healthy individuals from the Princeton University community with 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/hr). One participant was excluded from analysis because they moved their head between runs so much that their occipital lobe was partly shifted outside the field of view. The final sample consisted of 24 participants (15 women, mean age = 23 years). We previously reported some findings from this data set (Kok & Turk-Browne, 2018), though we performed additional analyses for present purposes that are reported here.

Experiment 2 enrolled 24 healthy individuals from the Yale University community with normal or corrected-to-normal vision (15 women, mean age = 24 years). Participants provided informed consent to a protocol approved by the Yale University Human Investigation Committee and were compensated ($20/hr). This is a new data set not previously reported.

Stimuli

Visual stimuli were generated using MATLAB (Mathworks; RRID:SCR_001622) and the Psychophysics Toolbox (Brainard, 1997; RRID:SCR_002881). In both experiments, stimuli were displayed on a rear-projection screen using a projector (1920 × 1068 resolution, 60 Hz refresh rate) against a uniform gray background. Participants viewed the stimuli through a mirror mounted on the head coil. Auditory cues consisted of three pure tones (440, 554, and 659 Hz; 80 msec per tone; 5-msec intervals), presented in ascending or descending pitch through headphones.

In Experiment 1, the visual stimuli were complex shapes defined by radio frequency components (RFCs; Drucker & Aguirre, 2009; Op de Beeck, Wagemans, & Vogels, 2001; Zahn & Roskies, 1972; Figure 1A–D). These stimuli were created by varying seven RFCs and were based on a subset of the stimuli used by Op de Beeck et al. (2001, see their Figure 1A). By varying the amplitude of three of the seven RFCs, a one-dimensional shape space was created. 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). Five shapes were chosen along this continuum to create a perceptually symmetrical set, centered on the third shape (for details, see Kok & Turk-Browne, 2018). To create slightly warped versions of the shapes, to enable a same/different discrimination task, a fourth RFC (the 3.18-Hz component) was modulated. The shape stimuli were presented

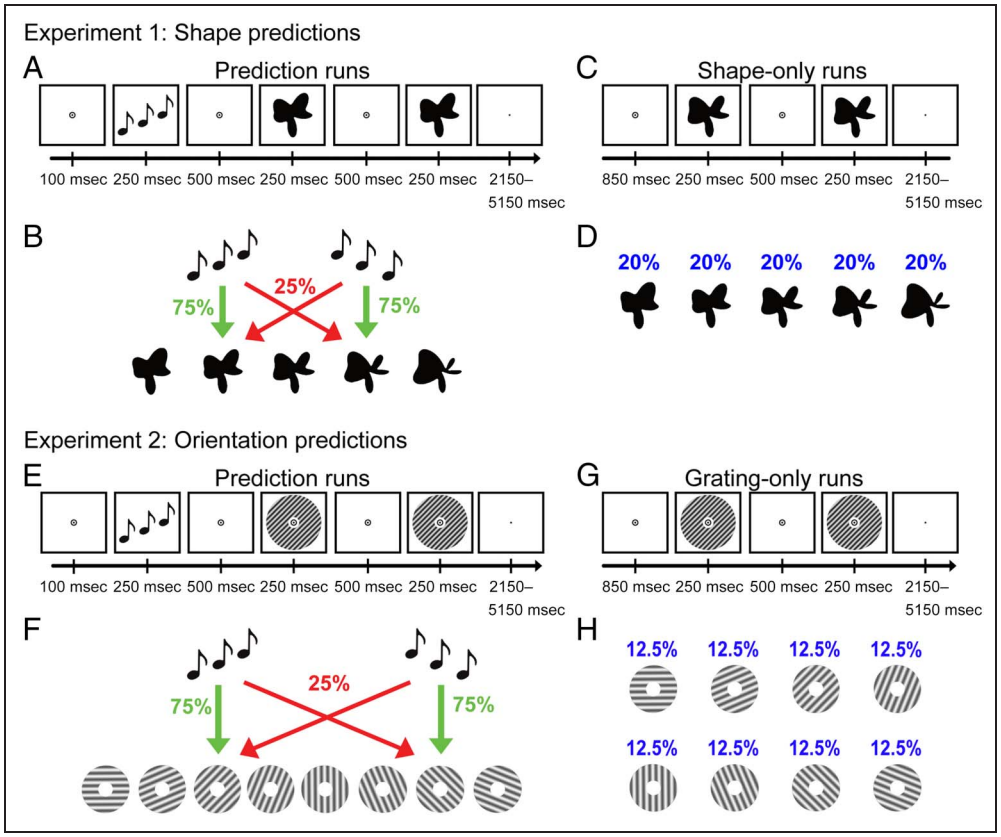

Figure 1. Experimental paradigms. (A) During prediction runs in Experiment 1, an auditory cue preceded the presentation of two consecutive shape stimuli, where the second shape was either identical to the first or slightly warped. Participants indicated whether the two shapes were the same or different. (B) The auditory cue (ascending vs. descending tones) predicted whether the first shape would be Shape 2 or Shape 4 (of five shapes). The cue was valid on 75% of trials, whereas in the other 25% of (invalid) trials the other, unpredicted shape was presented (prediction runs) or the shapes were omitted (omission runs). (C) During shape-only runs, no auditory cues were presented. As in the other runs, two shapes appeared sequentially and participants reported same or different. (D) All five shapes appeared with equal (20%) probability on trials of the shape-only runs. (E) The design of Experiment 2 was identical to Experiment 1, except that the shapes were replaced by oriented gratings. The second grating was either the same as the first or its phase was shifted slightly. (F) The auditory cue predicted whether the first grating would be rotated 45° or 135°, and this cue was valid 75% of the time. (G) During grating-only runs, no predictive auditory cues were presented. (H) All eight gratings were equally likely to appear (12.5%) during grating-only runs.

in black (subtending 4.5°), centered on a fixation bull's-eye.

In Experiment 2, visual stimuli consisted of grayscale luminance-defined sinusoidal gratings that were displayed in an annulus (outer diameter: 10°, inner diameter: 1°, spatial frequency: 1.5 cycles/°), surrounding a fixation bull's-eye (Figure 1E–H). Eight gratings were used to span the 180° orientation space, in equal steps of 22.5°. To enable a similar same/different discrimination task as in Experiment 1, we modulated the phase of the gratings.

Experimental Procedure

Each trial of Experiment 1 started with the presentation of a fixation bull's-eye (0.7°). During the "prediction" runs, an auditory cue (ascending or descending tones, 250 msec) was presented 100 msec after onset of the trial. After a 500-msec delay, two consecutive shape stimuli were presented for 250 msec each, separated by a 500-msec blank screen (Figure 1A). The auditory cue (ascending vs. descending tones) predicted whether the first shape on that trial would be Shape 2 or Shape 4, respectively (out of five shapes; Figure 1B). 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. During omission runs, the cues were also 75% valid, but on the remaining 25% of trials, no shape was presented at all, with only the fixation bull's-eye remaining on screen. All participants performed two prediction runs (128 trials, ~13 min per run) and two omission runs (128 trials, ~13 min per run), in interleaved ABBA fashion (order counterbalanced across participants). Halfway through the experiment, the contingencies between the auditory cues and the shapes were flipped (e.g., ascending tones were now followed by Shape 4 and descending by Shape 2). The order of the cue–shape mappings was counterbalanced across participants. Participants were trained on the cue–shape associations during two practice runs (112 trials total, ~8 min) in the scanner, one before the first prediction/omission run and one halfway through the experiment after the contingency reversal. During these practice runs, the auditory cue was 100% predictive of the identity of the first shape on that trial (e.g., ascending tones were always followed by Shape 2 and descending tones by Shape 4). The two practice runs took place while anatomical scans (see below) were acquired to make full use of scanner time. 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 orthogonal 3.18-Hz RFC component defining the shape by an amount much smaller than the differences between shape indices on the continuum defined over the three other varying components. This modulation could be either positive or negative (counterbalanced across conditions), and participants' task was to indicate whether the two shapes on a given trial were the same or different.

The design of Experiment 2 was identical to Experiment 1, except that the visual stimuli consisted of oriented gratings instead of complex shapes (Figure 1E–H). That is, Experiment 2 contained two prediction and two omission runs, but here, the auditory cues (rising and falling tones, 250 msec) predicted the orientation of an upcoming grating (45° or 135°, 250 msec). A second grating (250 msec) with the same orientation was presented after a 500-msec delay and was either identical or slightly phase shifted with respect to the first grating. As with the shape modulation for the same/different task, changes in phase were orthogonal and much smaller than the differences in orientation across stimuli. Participants' task was to indicate whether the two gratings were the same or different.

Finally, both experiments contained two additional runs in which no auditory cues were presented. Each trial started with the presentation of the fixation bull's-eye, a delay of 850 msec (to equate onset with runs containing the auditory cues), the first stimulus (250 msec), another delay of 500 msec, and the second stimulus (250 msec). The two visual stimuli (Experiment 1: complex shapes, Experiment 2: oriented gratings) were either identical or slightly different (Experiment 1: warped shape, Experiment 2: phase shift). Participants indicated whether the two visual stimuli were the same or different. In Experiment 1 (120 trials, ~13 min per run), each trial contained one of the five shapes with equal (20%) likelihood (Figure 1D). In Experiment 2 (128 trials, ~13 min per run), each trial contained one of the eight gratings with equal (12.5%) likelihood (Figure 1H). These "nonpredictive" runs were designed to be as similar as possible to the prediction/omission runs, save the absence of the predictive auditory cues. They were collected as the first and last runs of each session, and the data were used to train the neural decoding models (see below).

In both experiments, participants indicated their response using an MR-compatible button box. After the response interval ended (750 msec after disappearance of the second visual stimulus), the fixation bull's-eye was replaced by a single dot, signaling the end of the trial while still requiring participants to fixate. Also in both experiments, the magnitude of the difference between the two stimuli on a given trial (Experiment 1: shape warp, Experiment 2: phase offset) was determined by an adaptive staircasing procedure (Watson & Pelli, 1983), updated after each trial, to make the same/different task challenging (~75% correct) and comparable in difficulty across experiments. This staircase was implemented using Quest (psych.nyu.edu/pelli/software.html#quest), a Bayesian adaptive psychometric procedure that places each trial at the current most probably Bayesian estimate of the 75% accuracy threshold. In Experiment 1, this threshold was expressed as the logarithm of the difference in amplitude of the 3.18-Hz RFC component defining the shapes. In Experiment 2, it was the logarithm of the difference in phase of the two gratings. These small, just-detectable differences between the stimuli were thus updated on a trial-by-trial basis to compensate for potential task learning and fatigue effects over time. Separate staircases were run for trials containing valid and invalid cues, as well as for the nonpredictive runs, to equate task difficulty between conditions. That is, by running separate staircases for the different conditions, task difficulty was adjusted such that participants got approximately 75% correct in each condition. The staircases were kept running throughout the experiments. They were initialized at a value determined during an initial practice session 1–3 days 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 contingencies (valid trials only; 16 trials per contingency). Because participants were practicing both mappings equally, this session did not serve to train them on any particular cue–shape association. Rather, this session was intended to familiarize participants with the structure of trials and the nature of the experiment.

MRI Acquisition

For Experiment 1, structural and functional data were collected using a 3T Siemens Prisma scanner with a 64-channel head coil at the Princeton Neuroscience Institute. Functional images were acquired using a multiband EPI sequence (repetition time [TR] = 1000 msec, echo time [TE] = 32.6 msec, 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 most of the temporal and occipital lobes. Anatomical images were acquired using a T1-weighted MPRAGE sequence (TR = 2300 msec, TE = 2.27 msec, voxel size = 1 × 1 × 1 mm, 192 sagittal slices, 8° flip angle, GeneRalized Autocalibrating Partial Parallel Acquisition [GRAPPA] acceleration factor = 3). Two T2-weighted turbo spin-echo images (TR = 11,390 msec, TE = 90 msec, 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 for hippocampal segmentation. To correct for susceptibility distortions in the EPI, 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 msec, TE = 66 msec).

For Experiment 2, data were acquired on a 3T Siemens Prisma scanner with a 64-channel head coil at the Yale Magnetic Resonance Research Centre. Functional images were acquired using a multiband EPI sequence with virtually identical parameters to Experiment 1 (TR = 1000 msec, TE = 33.0 msec, 60 transversal slices, voxel size = 1.5 × 1.5 × 1.5 mm, 55° flip angle, multiband factor = 6), as was the pair of opposite phase-encode spin-echo volumes for distortion correction (TR = 8000 msec, TE = 66 msec). Anatomical images were similar to Experiment 1. T1-weighted images were acquired using an MPRAGE sequence (TR = 1800 msec, TE = 2.26 msec, voxel size = 1 × 1 × 1 mm, 208 sagittal slices, 8° flip angle, GRAPPA acceleration factor = 2). Two T2-weighted turbo spin-echo images were acquired (TR = 11,170 msec, TE = 93 msec, voxel size = 0.44 × 0.44 × 1.5 mm, 54 coronal slices, distance factor = 20%, 150° flip angle).

fMRI Preprocessing

Images for both experiments were preprocessed using FEAT 6 (fMRI Expert Analysis Tool), part of FSL 5 (fsl.fmrib.ox.ac.uk/fsl, Oxford Centre for Functional MRI of the Brain, RRID:SCR_002823; Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012). All analyses were performed in participants' native space. Using FSL's topup tool (Andersson, Skare, & Ashburner, 2003), susceptibility-induced distortions were determined on the basis of opposing-phase spin-echo volumes. This output was converted to radians per second and supplied to FEAT for B0 unwarping. The first six volumes of each run were discarded to allow T1 equilibration, and the remaining functional images for each run were spatially realigned to correct for head motion. These functional images were registered to each participant's T1 image using boundary-based registration and temporally high-pass filtered with a 128-sec period cutoff. No spatial smoothing was applied. Lastly, the two T2 images were coregistered and averaged, and the resulting image was registered to the T1 image through FLIRT (FMRIB's Linear Image Registration Tool).

ROIs

Our main focus was the hippocampus. Using the automatic segmentation of hippocampal subfields machine learning toolbox (Yushkevich et al., 2015) and a database of manual MTL segmentations from a separate set of 51 participants (Aly & Turk-Browne, 2016a, 2016b), hippocampal ROIs were defined based on each participant's T2 and T1 images for CA2-CA3-DG, CA1, and subiculum subfields. CA2, CA3, and DG were combined into a single ROI because these subfields are difficult to distinguish with fMRI. Results of the automated segmentation were visually inspected for each participant to ensure accuracy.

In visual cortex, ROIs were defined for V1, V2, and lateral occipital (LO) cortex in each participant's T1 image using Freesurfer (surfer.nmr.mgh.harvard.edu/; RRID:SCR_001847). To ensure that we were measuring responses in the retinotopic locations corresponding to our visual stimuli, we restricted the visual cortex ROIs to the 500 most active voxels during the nonpredictive runs. Because no clear retinotopic organization is present in the hippocampal ROIs, cross-validated feature selection was used instead (see below). All ROIs were collapsed over the left and right hemispheres, because we had no hypotheses regarding hemispheric differences.

fMRI Data Modeling

Functional data were modeled with general linear models using FILM (FMRIB's improved linear model). This included temporal autocorrelation correction and extended motion parameters (six standard parameters, plus their derivatives and their squares) as nuisance covariates.

For Experiment 1, we specified regressors for the conditions of interest: shape-only runs, five shapes; prediction runs, 2 shapes × 2 prediction conditions (valid vs. invalid); omission runs, 2 shapes × 2 omission conditions (presented vs. omitted). Delta functions were inserted at the onset of the first shape (or expected onset, for omissions) of each trial and convolved with a double-gamma hemodynamic response function (HRF). The same procedure was used for Experiment 2, with the following convolved regressors: grating only runs, eight orientations; prediction runs, 2 orientations × 2 prediction conditions (valid vs. invalid); omission runs, 2 orientations × 2 omission conditions (presented vs. omitted). For both experiments, we also included the temporal derivative of each regressor to accommodate variability in the onset of the response (Friston et al., 1998).

Additional finite impulse response (FIR) models were fit in both experiments to investigate the temporal evolution of shape and grating representations in visual cortex. This approach estimated the BOLD signal evoked by each condition of interest at 20 × 1 sec intervals. We trained the decoder on the FIR parameter estimates from the shape-only or grating-only runs, averaging over the time points spanning 4–7 sec (corresponding to the peak hemodynamic response). This decoder was then applied to the FIR parameter estimates from all of the time points in the prediction and omission runs. The amplitude and latency of this time-resolved shape/grating information was quantified by fitting a double gamma function and its temporal derivative to the decoder output.

Decoding Analysis

In both experiments, we used a forward modeling approach to reconstruct visual stimuli from the pattern of BOLD activity in a given brain region (Brouwer & Heeger, 2009). This approach has proven successful in reconstructing continuous stimulus features, such as hue (Brouwer & Heeger, 2009), orientation (Brouwer & Heeger, 2011),

and motion direction (Kok, Brouwer, Van Gerven, & De Lange, 2013).

In our case, the continuous dimensions of shape contour and grating orientation were modeled by a set of hypothetical channels, each with an idealized tuning curve (i.e., basis function). To decode shape in Experiment 1, the basis functions consisted of five hypothetical channels, each a halfwave-rectified sinusoid raised to the fifth power, spaced evenly such that they were centered on the 5 points in shape space that constituted the five shapes presented in the experiment (Figure 2A). To decode orientation in Experiment 2, the basis functions consisted of six halfwave-rectified sinusoids raised to the fifth power spaced evenly along the 180° orientation space (Figure 2D). The number of orientation basis functions was based on previous studies decoding circular stimulus features (Kok et al., 2013; Brouwer & Heeger, 2009, 2011). Note that reducing the number of basis functions in Experiment 2 to five, to match the number used in Experiment 1, yielded qualitatively identical results (data not shown). It was not possible to define six channels for Experiment 1, because the number of channels cannot exceed the number of exemplars (i.e., five). Finally, the orientation space, unlike the shape space, was circular, and therefore, the channels wrapped around from 180° to 0°.

Other than these subtle differences in the definitions of the basis functions, the forward modeling approach was identical in Experiments 1 and 2. In the first "training" stage, BOLD activity patterns obtained from the shape-only or grating-only runs were used to estimate the weights of these channels 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 or the eight orientations in the nonpredictive runs). The matrix of estimated BOLD response

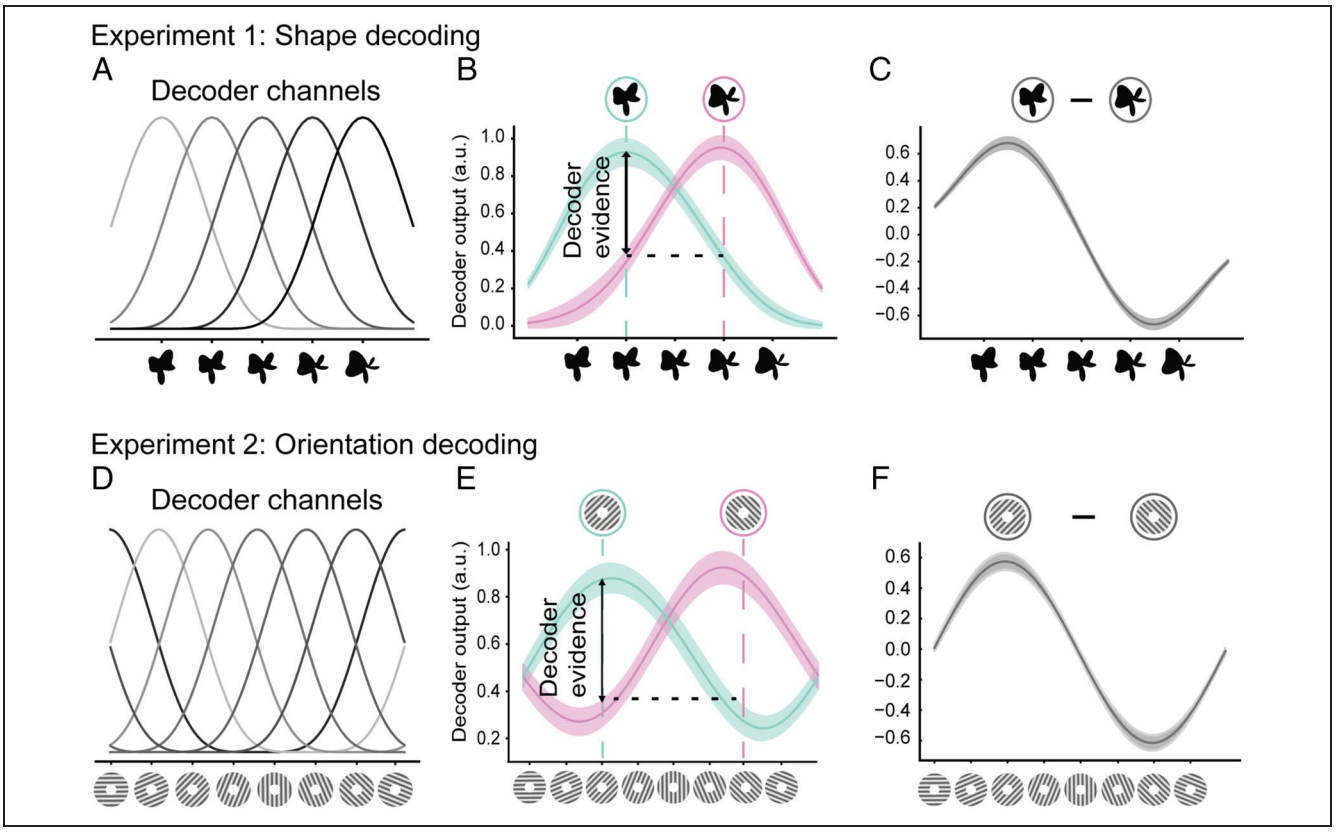

Figure 2. Illustration of the decoding methods. (A) We used a forward modeling approach to reconstruct shapes from the pattern of BOLD activity. Shape selectivity was characterized by five hypothetical channels, each with an idealized shape tuning curve. BOLD patterns obtained from the shape-only runs 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 and omission 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 evidence across shape space. Decoding performance was quantified by subtracting the evidence at the presented shape (e.g., Shape 2) from the evidence at the non-presented shape (e.g., Shape 4). (C) Finally, solely for the purpose of visualising the shape-specific information, we collapsed across the presented shapes by subtracting the neural evidence for Shape 4 from that for Shape 2, thereby removing any non-shape-specific BOLD signals. (D) An identical forward modeling approach was used in Experiment 2, except that orientation selectivity was characterized by six hypothetical channels that wrapped around the circular orientation space. (E) Decoding performance was quantified by subtracting the neural evidence at the presented orientation (e.g., 45°) from the evidence at the non-presented orientation (e.g., 135°). (F) As above, we collapsed across the presented gratings by subtracting the neural evidence for the presented (e.g., 45°) from the nonpresented (e.g., 135°) orientation, for visualization purposes. Shaded bands in B, C, E, and F indicate *SEM*.

amplitudes for the different stimuli ($\mathbf{B_{train}}$, $m \times n$) was related to the matrix of hypothetical channel outputs ($\mathbf{C_{train}}$, $k \times n$) by a weight matrix ($\mathbf{W}$, $m \times k$):

$$\mathbf{B_{train}} = \mathbf{W}\mathbf{C_{train}} \quad (1)$$

The least-squares estimate of this weight matrix $\mathbf{W}$ was estimated using linear regression:

$$\hat{\mathbf{W}} = \mathbf{B_{train}}\mathbf{C}_{\mathbf{train}}^{\mathrm{T}}\left(\mathbf{C_{train}}\mathbf{C}_{\mathbf{train}}^{\mathrm{T}}\right)^{-1} \quad (2)$$

These weights reflected the relative contribution of the hypothetical channels in the forward model to the observed response amplitude of each voxel. Using these weights, the second stage of analysis reconstructed the channel outputs associated with the test patterns of activity across voxels evoked by the stimuli in the main experiment (i.e., the prediction and omission runs; $\mathbf{B_{test}}$), again using linear regression. This step transformed each vector of n voxel responses (parameter estimates per condition) into a vector of k channel responses. These channel responses ($\mathbf{C_{test}}$) were estimated using the learned weights ($\mathbf{W}$):

$$\hat{\mathbf{C}}_{\mathbf{test}} = \left(\hat{\mathbf{W}}^{\mathrm{T}}\hat{\mathbf{W}}\right)^{1}\hat{\mathbf{W}}^{\mathrm{T}}\mathbf{B_{test}} \quad (3)$$

The channel outputs were used to compute a weighted average of the basis functions, reflecting neural evidence over the shape or orientation dimension (Figure 2B, E). During the prediction and omission runs of Experiment 1, only Shapes 2 and 4 were presented (Figure 1B). Thus, four neural evidence curves were obtained for these runs: two shapes by two prediction/omission conditions (valid vs. invalid/presented vs. omitted). We collapsed across the presented shape by subtracting the neural evidence for Shape 4 from that for Shape 2, thereby subtracting out any nonshape-specific BOLD signals (Figure 2C). Analogously, in Experiment 2, we subtracted the neural evidence for 135° gratings from that for 45° gratings (Figure 2F).

For statistical testing, scalar decoding performance values were calculated on the basis of decoded neural evidence. For Experiment 1, decoding performance during the prediction/omission runs was quantified by subtracting the neural evidence for the presented shape (e.g., Shape 2) from that of the nonpresented shape (e.g., Shape 4). For Experiment 2, decoding performance was quantified by subtracting the neural evidence for the presented orientation (e.g., 45°) from that of the nonpresented orientation (e.g., 135°). (Note that, for the omission trials, there was no presented shape or orientation, and so we conditioned decoding performance on the "expected" shape or orientation.) For each participant in the two experiments, this procedure led to a measure of decoding performance for validly and invalidly predicted stimuli (shapes or gratings), respectively. This allowed us to quantify evidence for the stimuli as presented on the screen (by averaging evidence for validly and invalidly predicted stimuli) and evidence for the cued stimuli (by averaging [1 – evidence] for the invalidly predicted stimuli with evidence for the validly predicted stimuli). Finally, we calculated decoding performance for predicted but omitted stimuli (shapes and gratings, respectively). These measures were statistically tested at the group level using mixed-design ANOVAs and independent samples t tests (see Experimental Design and Statistical Analysis section).

For all ROIs, voxel selection was based on data from the shape- and grating-only runs, in which no predictions were present to ensure voxel selection was independent of the data in which we tested our effects of interest (i.e., the prediction and omission runs). In visual cortex ROIs, we selected the 500 most active voxels during the shape- and grating-only runs. However, the hippocampus does not show a clear evoked response to visual stimuli, as defined by a lack of significant fit of a regressor of stimulus onset times convolved with a canonical hemodynamic response to the mean hippocampal time course. Therefore, we applied a different method of voxel selection for hippocampal ROIs. Voxels were first sorted by their informativeness, that is, how different the weights for the forward model channels were from each other, as indexed by the standard deviation of the weights. Second, the number of voxels to include was determined by selecting between 10% and 100% of the voxels, increasing in 10% increments. We then trained and tested the model on these voxels within the shape- and grating-only runs (trained on one run and tested on the other). For all iterations, decoding performance on Shapes 2 and 4 (Experiment 1) or 45° and 135° orientations (Experiment 2) was quantified as described above, and we selected the number of voxels that yielded the highest performance (group average: Experiment 1, 1536 of 3383 voxels; Experiment 2, 1369 of 3229 voxels). We also labeled the selected hippocampus voxels based on their subfield from segmentation (group average: Experiment 1, 436 voxels in CA1; 572 voxels in CA2-CA3-DG; 425 voxels in subiculum; Experiment 2, 394 voxels in CA1; 490 voxels in CA2-CA3-DG; 380 voxels in subiculum).

For the hippocampal ROIs and searchlight analyses, the input to the forward model consisted of parameter estimates from a voxelwise general linear model that fit the amplitude of the BOLD response using regressors convolved with a double-gamma HRF. However, for the visual cortex ROI analyses, we supplied parameter estimates from a data-driven FIR model that made no assumptions about the timing or shape of BOLD responses. In this analysis, the amplitude and latency of decoded shape/orientation information was quantified by fitting a double gamma function and its temporal derivative to the decoder output.

Searchlight

A searchlight approach was used to explore the specificity of predicted shape and orientation representations, as

well as significant differences between the two, within the field of view of our functional scans (most of occipital and temporal and part of parietal and frontal cortex). In both experiments, a spherical searchlight with a radius of 5 voxels (7.5 mm) was passed over all functional voxels, using the searchlight function implemented in the Brain Imaging Analysis Kit (BrainIAK, brainiak.org, RRID: SCR_014824). In each searchlight for each participant, we performed shape/orientation decoding in the same manner as in the ROIs, which yielded maps of decoding evidence for the predicted shapes (Experiment 1) and orientations (Experiment 2).

Each participant's output volumes were registered to the Montreal Neurological Institute (MNI) 152 standard template for group analysis. 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). Group-level nonparametric permutation tests were applied to these searchlight maps using FSL randomise (Winkler, Ridgway, Webster, Smith, & Nichols, 2014), correcting for multiple comparisons using threshold-free cluster enhancement (Smith & Nichols, 2009). To determine where in the brain orientation and shape prediction signals differed from one another, we conducted a two-sample t test comparing predicted shape evidence to predicted orientation evidence at $p < .05$ (two-sided). This test yielded one large cluster, so we identified local maxima within the cluster by reducing the critical p value to .005. Follow-up one-sample t tests were used to explore the two experiments separately at $p < .05$ (one-sided).

Experimental Design and Statistical Analysis

The aim of the current study was to compare differences in neural representations evoked by predictions of complex shapes and grating orientations, respectively. The decoded neural representations for the complex shapes (Experiment 1) were separately reported on in a previous publication (Kok & Turk-Browne, 2018), and the decoding results of that study, in hippocampus and visual cortex, were retained for this study. The novelty of this study lies in directly comparing these representations to those evoked by simple grating orientations (Experiment 2). Additionally, we here report neural representations of predicted-but-omitted stimuli in hippocampus and visual cortex, for both complex shapes and simple gratings.

For hippocampal ROIs, we quantified decoding evidence for validly and invalidly predicted stimuli (shapes and orientations, respectively) and subjected these measures to mixed-design two-way ANOVAs with Prediction Validity (valid vs. invalid) as the within-group factor and Stimulus Type (shape vs. orientation) as the between-group factor. Significant interactions between these two factors indicate differential effects of predictions depending on stimulus type. Follow-up tests of the effect of prediction validity within experiments were conducted using paired-sample t tests.

We also quantified decoding evidence for predicted-but-omitted stimuli (shapes and orientations, respectively) as a measure of "pure" prediction effects in the absence of actually presented visual stimuli. This analysis focused on the neural activity evoked by the trials in which an auditory cue predicted a specific stimulus (shape or orientation), but the screen remained empty (except for the fixation bull's-eye). Decoding evidence for these predicted-but-omitted shapes and orientations was quantified per participant, and the effect of stimulus type (shape vs. orientation) was tested by submitting these measures to independent-samples t tests. Effects of predicted-but-omitted stimuli within experiment were tested using one-sample t tests.

For visual cortex ROIs, we quantified decoder evidence in a time-resolved manner (see fMRI Data Modeling section and Decoding Analysis section above) to be able to investigate influences of prediction on both the amplitude and the latency of neural responses. This time-resolved analysis was motivated by our previous finding of prediction modulating the latency of BOLD signals in visual cortex (Kok & Turk-Browne, 2018). The amplitudes and latencies of decoding signals were quantified by fitting a canonical HRF and its temporal derivative, respectively, to the decoder evidence time courses for validly and invalidly predicted stimuli (shapes or orientations). These amplitude and latency estimates were subjected to mixed-design two-way ANOVAs with Prediction Validity (valid vs. invalid) as the within-group factor and Stimulus Type (shape vs. orientation) as the between-group factor. Decoding amplitude and latency for predicted-but-omitted stimuli were compared between stimulus types (shapes vs. orientations) using independent-samples t tests.

RESULTS

Participants were exposed to auditory cues that predicted either which complex shape was likely to be presented (Experiment 1, Figure 1A–D) or the likely orientation of an upcoming grating stimulus (Experiment 2, Figure 1E–H). In both experiments, two stimuli were presented on each trial, which were either identical or slightly different from one another (Experiment 1: second shape slightly warped, Experiment 2: second grating slightly different phase). Participants were asked to report whether the two stimuli on any given trial were the same or different.

Behavioral Results

Participants were able to discriminate small differences in both complex shapes (36.9 ± 2.3% modulation of the 3.18-Hz radial frequency component, mean ± *SEM*) and simple gratings (2.7 ± 0.2 radians phase difference, mean ± *SEM*) during the visual stimuli only runs. This

was also the case during the prediction runs for both complex shapes (valid trials, 31.6 ± 2.5%; invalid trials, 33.2 ± 2.9%; no significant difference in discrimination threshold between valid and invalid trials, $t(23) = 1.00$, $p = .32$) and simple gratings (valid trials, 2.4 ± 0.2 radians; invalid trials, 2.4 ± 0.2 radians; no significant difference, $t(23) = 0.36$, $p = .72$). For Experiment 1, accuracy and RTs did not differ between valid trials (accuracy, 70.6 ± 1.2%; RT, 575 ± 16 msec) and invalid trials (accuracy, 68.8 ± 1.5%; RT, 573 ± 18 msec; both ps > .20). In Experiment 2, participants were slightly more accurate for valid (75.2 ± 1.3%) than invalid (72.0 ± 1.5%) trials, $t(23) = 2.25$, $p = .03$. Note that this difference in accuracy indicates that the staircase procedure did not perfectly equate task performance for valid versus invalid trials in Experiment 2. We address whether task difficulty affected hippocampus representations below (see Control Analyses section). RTs did not differ significantly between conditions (valid: 646 ± 13 msec, invalid: 646 ± 14 msec, $p = .93$).

Hippocampus

Predictions about complex shapes and grating orientations led to strikingly different responses in the hippocampus (interaction between Cue Validity and Stimulus Type; $F(1, 46) = 9.14$, $p = .0041$; no main effects, ps > .3; Figure 3). In both experiments, we quantified evidence for the stimuli as presented on the screen (by averaging evidence for the stimulus whether validly or invalidly predicted) and evidence for the predicted stimuli (by averaging evidence for the stimulus when validly predicted with [1 – evidence] for the stimulus when invalidly predicted). In Experiment 1, the pattern of activity in the hippocampus contained a representation of the shape that was predicted by the auditory cue, $t(23) = 2.86$, $p = .0089$ (Figure 3A) but was unaffected by the shape that was actually presented on screen, $t(23) = 0.54$, $p = .59$. However, the situation was strikingly different for orientation in Experiment 2, where the pattern of activity in the hippocampus did not contain a significant representation of the orientation that was predicted by the cue, $t(23) = -1.59$, $p = .13$, or the presented orientation, $t(23) = 1.35$, $p = .19$ (Figure 3B). In short, there was a difference between shape and orientation prediction signals in the hippocampus, $F(1, 46) = 9.14$, $p = .0041$ (Figure 3C), driven by a positive shape prediction signal in Experiment 1 and a numerically negative orientation prediction signal in Experiment 2.

To interrogate the circuitry of these prediction signals further, we applied an automated segmentation method to define ROIs for the anatomical subfields of the hippocampus. Specifically, we segmented the hippocampus into CA1, CA2-CA3-DG, and subiculum. As in the hippocampus as a whole, representations of predicted shapes and orientations were strikingly different in CA2-CA3-DG (interaction between Cue Validity and Stimulus Type; $F(1, 46) = 9.40$, $p = .0036$; no main effects, ps > .1; Figure 4C). Where Experiment 1 showed a trend toward evoking a "positive" representation of the predicted shape in this subregion, $t(23) = 2.04$, $p = .053$ (Figure 4A), orientation predictions were "negatively" represented, $t(23) = -2.29$, $p = .031$ (Figure 4B).

To inspect these results more closely, consider the responses evoked by valid and invalid predictions in both experiments. In Experiment 1, hippocampal representations were completely determined by the predicted shape. That is, when Shape 2 was predicted and presented,

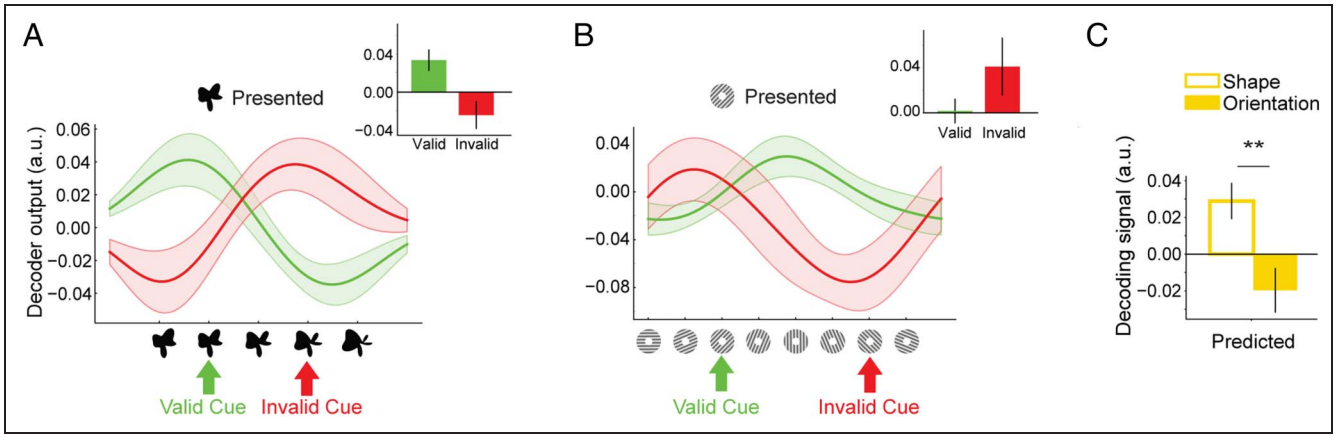

Figure 3. Stimulus reconstructions in hippocampus. (A) Decoder output of a forward model trained on shape-only runs, applied separately to validly (green) and invalidly (red) predicted shapes in the prediction runs. Collapsed across trials in which Shapes 2 and 4, respectively, were presented (see Figure 2C). The inset depicts quantified decoding evidence (see Figure 2B) for validly and invalidly predicted shapes. (B) Decoder output of a forward model trained on grating-only runs, applied separately to validly (green) and invalidly (red) predicted gratings in the prediction runs. Collapsed across trials in which 45° and 135° oriented gratings, respectively, were presented (see Figure 2F). The inset depicts quantified decoding evidence (see Figure 2E) for validly and invalidly predicted orientations. (C) Decoding evidence for predicted shapes (outlined bar) and predicted orientations (filled bar), quantified by averaging (1 − evidence) for the invalidly predicted stimuli with evidence for the validly predicted stimuli. **$p < .01$. Shaded bands and error bars indicate *SEM*.

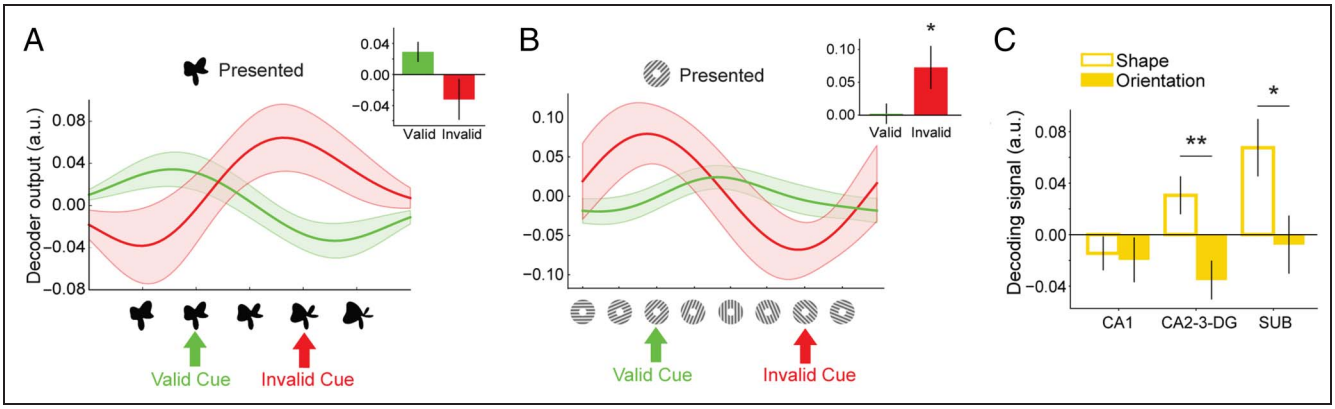

Figure 4. Stimulus reconstructions in CA2-CA3-DG. (A) Decoder output of a forward model trained on shape-only runs, applied separately to validly (green) and invalidly (red) predicted shapes in the prediction runs. Collapsed across trials in which Shapes 2 and 4, respectively, were presented (see Figure 2C). The inset depicts quantified decoding evidence (see Figure 2B) for validly and invalidly predicted shapes. (B) Decoder output of a forward model trained on grating-only runs, applied separately to validly (green) and invalidly (red) predicted gratings in the prediction runs. Collapsed across trials in which 45° and 135° oriented gratings, respectively, were presented (see Figure 2F). The inset depicts quantified decoding evidence (see Figure 2E) for validly and invalidly predicted orientations. (C) Decoding evidence for predicted shapes (outlined bars) and predicted orientations (filled bars), quantified by averaging (1 − evidence) for the invalidly predicted stimuli with evidence for the validly predicted stimuli, across hippocampal subfields. *$p < .05$, **$p < .01$. Shaded bands and error bars indicate *SEM*.

hippocampal patterns represented Shape 2, and when Shape 4 was predicted but Shape 2 was presented, hippocampus solely represented Shape 4 (Figures 3A and 4A). In Experiment 2, when grating orientation predictions were valid, activity patterns contained no evidence for any orientation, neither the predicted (and presented) orientation, nor the unpredicted orientation (hippocampus: $t(23) = 0.14$, $p = .89$; CA2-CA3-DG: $t(23) = 0.13$, $p = .89$; Figures 3B and 4B). When orientation predictions were invalid, however, activity patterns reflected the (unexpectedly) presented orientation, most clearly in CA2-CA3-DG ($t(23) = 2.18$, $p = .04$; hippocampus: $t(23) = 1.57$, $p = .13$). In other words, only unexpectedly presented grating orientations were represented in CA2-CA3-DG, reminiscent of a prediction error type signal.

In the subiculum, predicted shapes and orientations evoked distinct representations as well (interaction between Cue Validity and Stimulus Type; $F(1, 46) = 5.40$, $p = .025$; no main effects, ps $> .05$; Figure 4C). As in hippocampus as a whole, Experiment 1 revealed that shape representations in subiculum were dominated by the predicted shape, $t(23) = 2.97, p = .0069$, but not the presented shape, $t(23) = -0.54, p = .59$. In Experiment 2, on the other hand, neither predicted, $t(23) = -0.34$, $p = .74$, nor presented, $t(23) = 0.52$, $p = .61$, orientations were represented. That is, subiculum activity patterns did not contain information about grating orientation in any of the conditions. Finally, in CA1, there were no significant effects of prediction on decoding evidence (no main effects of Cue Validity or Stimulus Type, nor an interaction, all ps $> .1$).

These subfield results for valid versus invalid predictions were largely mimicked by representations evoked in omission trials. Predicted-but-omitted shapes affected activity patterns in CA2-CA3-DG strikingly differently than predicted-but-omitted orientations, $t(46) = 2.57$, $p = .013$ (Figure 5). This difference was in the same direction

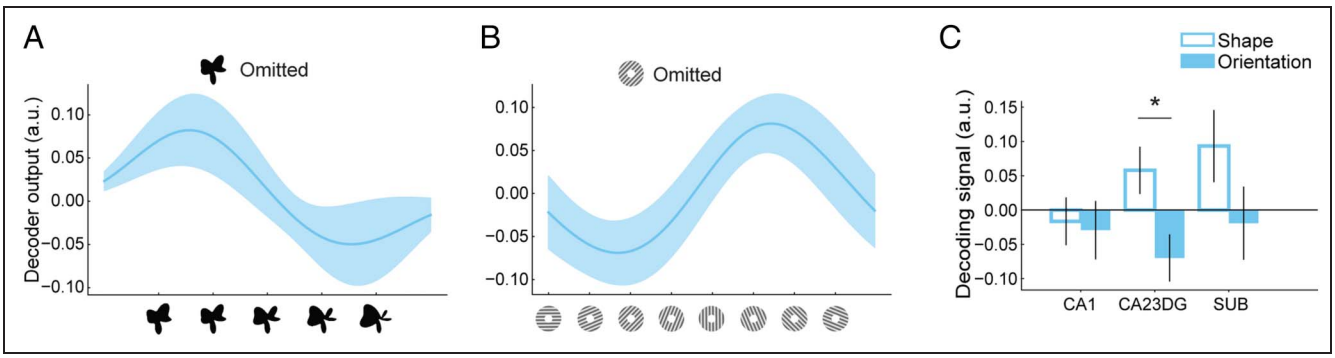

Figure 5. Reconstruction of predicted-but-omitted stimuli in CA2-CA3-DG. (A) Decoder output of a forward model trained on shape-only runs and applied to predicted-but-omitted shapes in the omission runs. Collapsed across trials in which Shapes 2 and 4, respectively, were predicted (see Figure 2C). (B) Decoder output of a forward model trained on grating-only runs, applied to predicted-but-omitted orientations in the omission runs. Collapsed across trials in which 45° and 135° oriented gratings, respectively, were predicted (see Figure 2F). (C) Decoding evidence for predicted-but-omitted shapes (outlined bars) and predicted-but-omitted orientations (filled bars), across hippocampal subfields. *$p < .05$. Shaded bands and error bars indicate *SEM*.

as above, with shape prediction signals being more positive than orientation prediction signals. Note that the positive shape prediction signals, $t(23) = 1.64$, $p = .11$, and negative orientation prediction signals, $t(23) = -2.00$, $p = .058$, were not significant in isolation. Representations of expected but omitted stimuli did not significantly affect subiculum (shapes: $t(23) = 1.73$, $p = .097$; orientations: $t(23) = -0.35$, $p = .73$; difference: $t(46) = 1.47$, $p = .15$) or CA1 (shapes: $t(23) = -0.46$, $p = .65$; orientations: $t(23) = -0.67$, $p = .51$; difference: $t(46) = 0.23$, $p = .82$).

Does the prediction error-like response to invalidly predicted orientations in CA2-CA3-DG depend on the appearance of an unexpected orientation that actively violates the prediction? We compared this response with the negative evidence for the cued orientation on omission trials in which with the prediction fails to materialize in time but without a violating stimulus. There was no significant difference between these two trial types (hippocampus: $t(23) = -0.24$, $p = .81$; CA2-CA3-DG: $t(23) = -0.07$, $p = .95$). In summary, the validity of orientation predictions influences hippocampal responses, because valid outcomes are cancelled out (Figures 3B and 4B), but the type of invalid outcome (unpredicted orientation or omission) does not seem to influence hippocampal responses.

Overall, the results for omission trials in the hippocampus resembled those for predicted stimuli when comparing valid and invalid trials. However, the effects were weaker statistically, perhaps because of lower signal-to-noise ratio on trials without any visual stimulus or because of a qualitative difference between these conditions. An example of the latter could be that the omission of expected stimuli may trigger different cognitive processes than validly and invalidly cued trials. The absence of any visual stimulus 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 decision-making and response selection.

Visual Cortex

In visual cortex ROIs, we applied the decoding analysis in a time-resolved manner and characterized the time courses of the decoding signal by fitting a canonical (double-gamma) 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 (Henson, Price, Rugg, Turner, & Friston, 2002; Friston et al., 1998).

Interestingly, the temporal evolution of stimulus representations in visual cortex was strongly affected by the auditory prediction cues. Valid predictions about complex shapes and grating orientations led to similar facilitation in visual cortex, across the cortical hierarchy (Figure 6). Specifically, when predictions were invalid, there was a delay in the decoding signal relative to when they were valid (main effect of validity on temporal derivative, V1: $F(1, 46) = 19.85$, $p = .000053$; V2: $F(1, 46) = 13.90$, $p = .00053$; LO: $F(1, 46) = 7.97$, $p = .0070$). This effect did not vary by stimulus type (no interaction between Validity and Stimulus Type, V1: $F(1, 46) = 0.49$, $p = .49$; V2: $F(1, 46) = 1.02$, $p = .32$; LO: $F(1, 46) = 0.04$, $p = .84$) and was present for both shapes (V1: $t(23) = 3.40$, $p = .0024$; V2: $t(23) = 3.06$, $p = .0056$; marginal effect in LO: $t(23) = 1.96$, $p = .062$) and orientations (V1: $t(23) = 2.87$, $p = .0086$; V2: $t(23) = 2.15$, $p = .042$; marginal effect in LO: $t(23) = 2.06$, $p = .051$). The peak of the decoding signal was significantly lower for invalidly predicted stimuli than for validly predicted stimuli in V1 (main effect of Validity on canonical HRF), $F(1, 46) = 7.97$, $p = .0070$, but not in V2, $F(1, 46) = 1.98$, $p = .17$, or LO, $F(1, 46) = 0.02$, $p = .90$. This effect in V1 did not depend significantly on stimulus type either (no interaction between Validity and Stimulus Type), $F(1, 46) = 1.04$, $p = .31$; however, in isolation, shapes were significant, $t(23) = 2.73$, $p = .012$, but not orientations, $t(23) = 1.27$, $p = .22$. In short, there were no significant differences in how predictions affected either the peak or the latency of the decoded stimulus signals between the two experiments, consistent with the possibility that shape and orientation expectations modulate visual cortex similarly.

During the omission runs, the 25% nonvalid trials did not involve the presentation of the unpredicted stimulus, but rather no visual stimulus at all. To investigate whether the BOLD response in visual cortex reflected the predicted-but-omitted stimuli, we inspected the fit of a canonical HRF to the decoding time course for these trials. This revealed that the orientation of expected but omitted gratings was successfully reconstructed from BOLD patterns in V1, $t(23) = 2.28$, $p = .032$, but not in V2, $t(23) = 1.37$, $p = .18$, or LO, $t(23) = 0.59$, $p = .56$, replicating Kok, Failing, and De Lange (2014). However, expected but omitted shapes could not be reconstructed from BOLD patterns in any region (V1: $t(23) = 0.64$, $p = .53$; V2: $t(23) = -1.33$, $p = .20$; LO: $t(23) = -0.71$, $p = .49$). Note that there were no reliable differences in decoding performance between omitted gratings and omitted shapes (V1: $t(46) = 1.06$, $p = .29$; V2: $t(46) = 1.91$, $p = .062$; LO: $t(46) = 0.92$, $p = .36$).

Searchlight Results

Our primary focus in this study was on the hippocampus, but to explore which other brain regions are involved in content-sensitive predictions, we performed searchlight analyses in the field of view of our functional scans (most of occipital and temporal and part of parietal and frontal cortex). This analysis revealed distinct representations of shape and orientation predictions in bilateral hippocampus, anterior occipital cortex, cerebellum, left inferior frontal gyrus (IFG), and left middle temporal gyrus, as

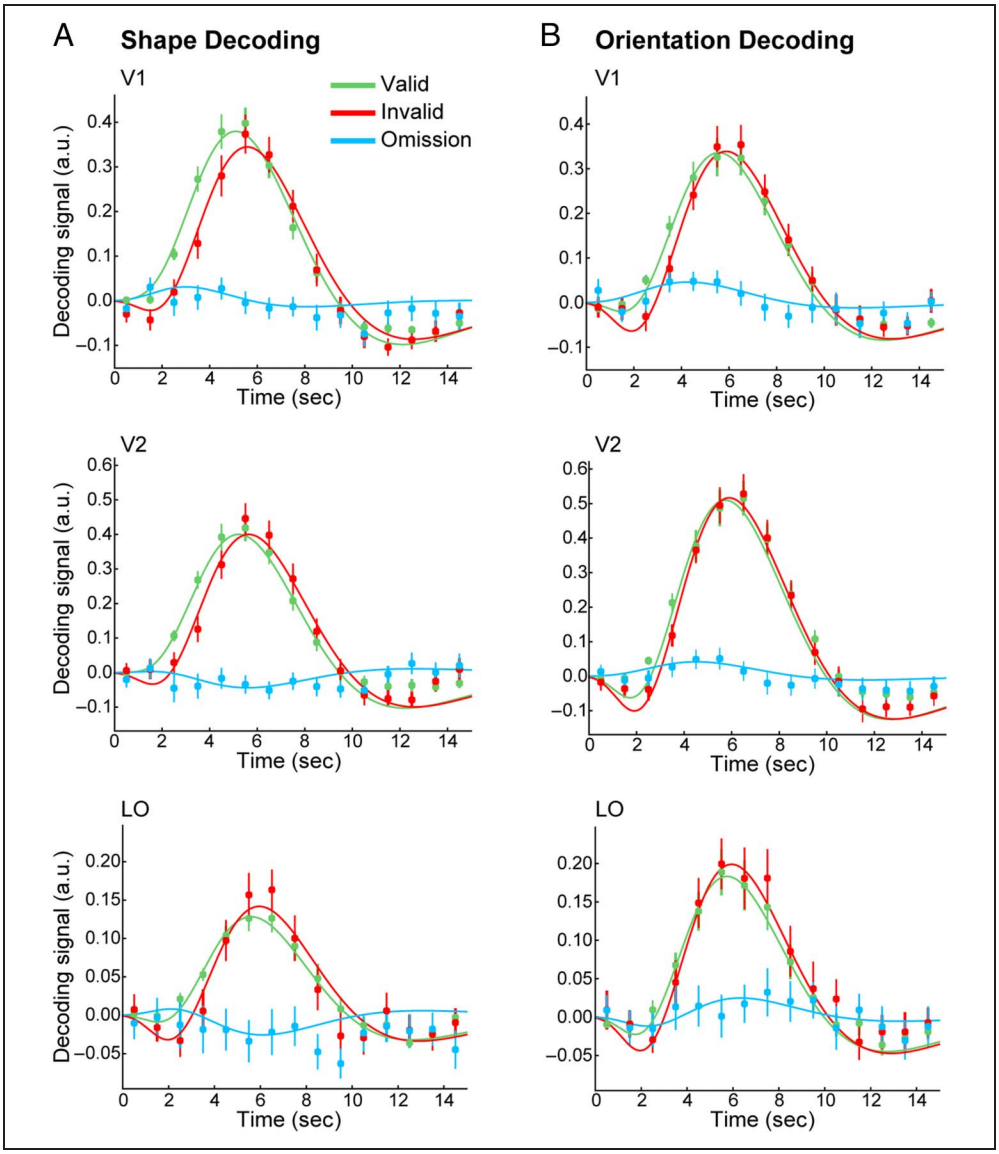

Figure 6. Time-resolved stimulus reconstructions in visual cortex. (A) Decoding evidence from a forward model trained on shape-only runs and applied in a time-resolved manner to validly (green) and invalidly (red) predicted shapes, as well as predicted-but-omitted shapes (blue). Individual data points reflect decoding evidence obtained for each time point from FIR parameter estimates; solid lines reflect fit with canonical HRF and its temporal derivative. (B) Decoding evidence from a forward model trained on grating-only runs and applied in a time-resolved manner to validly (green) and invalidly (red) predicted orientations, as well as predicted-but-omitted orientations (blue). Individual data points reflect decoding evidence obtained for each time point from FIR parameter estimates; solid lines reflect fit with canonical HRF and its temporal derivative. Error bars indicate *SEM*.

well as a few smaller clusters elsewhere (Table 1). Separate searchlight analyses for the two experiments revealed positive representations of the predicted shape in the hippocampus, anterior occipital cortex, and a few smaller clusters, and negative orientation predictions in left IFG, anterior occipital cortex, and middle temporal gyrus (Table 1). The reverse contrasts (i.e., negative shape predictions and positive orientation predictions) did not reveal any significant clusters.

Control Analyses

Can the differences between shape and orientation experiments be explained by the fact that the two gratings associated with auditory cues were easier to distinguish than the two shapes associated with auditory cues (or vice versa)? A difference in perceptual distance between the two shapes versus the two gratings may cause a different balance between pattern completion and pattern separation mechanisms in the hippocampus, thereby complicating our results. To examine this possibility, we split the participants in both experiments into two groups, depending on how well the two gratings or shapes could be decoded from V1, in the visual stimuli only runs. We then investigated whether hippocampal evidence for the predicted shape/orientation differed between high versus low V1 decoders, using a two-way ANOVA. If the difference between the two experiments was driven by stimulus discriminability, we would expect to see a main effect of High versus Low Decoders. However, we found a strong main effect of Experiment, $F(1, 44) = 9.79$, $p = .0031$; no effect of High versus Low Decoders, $F(1, 44) = 1.24$, $p = .27$; and no interaction between the two, $F(1, 44) = 3.62$, $p = .064$. The trend toward an interaction reflected the fact that the positive prediction signal for shapes and the negative prediction signal for orientations tended to be stronger for the participants with "worse" V1 stimulus decoding. We are

Table 1. Searchlight Results

Cluster Size (Voxels)	*Anatomical Region*	*Hemisphere*	*Peak p*	*Coordinates (x y z)*		
	Predicted Shape > Predicted Orientation Decoding					
16132	Calcarine sulcus	Left	.0018	−18	−56	4
		Right	.0036	12	−80	4
	Lingual gyrus	Left	.0024	−6	−72	−2
		Right	.0028	24	−64	2
	Cuneus	Left	.0036	−14	−80	18
		Right	.0024	14	−70	22
	Hippocampus	Left	.0036	−22	−19	−15
		Right	.0024	26	−20	−18
	Parahippocampal gyrus	Left	.0036	−26	−24	−18
		Right	.0036	22	−36	−10
	Cerebellum	Left	.0036	−22	−62	−22
		Right	.0028	8	−68	−12
	Pallidum	Left	.0018	−14	4	8
	Thalamus	Right	.0024	16	−24	0
	Inferior frontal gyrus	Left	.0028	−46	24	16
449	Middle temporal gyrus	Left	.017	−70	−36	−12
58	Hippocampus	Right	.018	24	−32	22
17	Insula	Left	.024	−38	−8	2
10	Superior temporal gyrus	Left	.023	−42	−14	−4
	Predicted Shape Decoding > 0					
180	Calcarine sulcus	Right	.016	14	−70	20
63	Hippocampus	Right	.026	24	−18	16
27	Middle cingulate	Right	.028	2	8	40
7	Caudate	Left	.028	−16	4	8
5	Cerebellum	Left	.044	−26	−62	22
	Predicted Orientation Decoding < 0					
393	Inferior frontal gyrus	Left	.0028	−46	24	16
252	Calcarine sulcus	Left	.0024	−24	−56	4
30	Middle temporal gyrus	Left	.020	−66	−36	−6

All p values are corrected for multiple comparisons. Coordinates reflect local maxima of significant clusters in MNI space. Local maxima within the largest cluster were identified by reducing the critical p value to .005. No clusters were obtained by reversing the sign of the comparisons.

hesitant to interpret a marginal effect, but this could potentially reflect the need for stronger prediction in participants who had trouble disambiguating the stimuli. Regardless, this is in the opposite direction than hypothesized above and thus does not provide evidence for an alternative explanation of content-sensitive hippocampal effects based on stimulus discriminability. In an additional control analysis, we compared decoding performance in the hippocampus for shapes and orientation in the stimulus-only run and found no significance

difference in decodability between stimulus types, $t(46) = 1.62$, $p = .11$. This suggests that discriminability per se is not a sufficient explanation for the dissociation reported here.

Can the differences between hippocampal representations of shape and orientation predictions be explained by orientation being a circular space, while the shapes were sampled from a noncircular space? Specifically, it may be hypothesized that the two orientations presented in the prediction runs evoke opposing (i.e., negatively correlated) patterns of activity in hippocampus because they are at opposite points of a circular space, whereas the two shapes are encoded in orthogonal (i.e., noncorrelated) patterns. Such a qualitative difference in the way the two stimulus spaces are encoded might affect how predictions about these stimuli are encoded as well. We investigated this by correlating the hippocampal patterns evoked by these stimuli in different stimulus-only runs, both for the same stimuli (e.g., Shape 2 in the first run with Shape 2 in the last run) and for different stimuli (e.g., Shape 2 in the first run with Shape 4 in the last run). These correlations were calculated within participants and assessed for reliability versus 0 at the group level. First, we found, as expected, a modest positive correlation for stimuli that were the same, both for shapes (Shape 2 with Shape 2, and Shape 4 with Shape 4; mean $r = .0269$, $t(23) = 2.22$, $p = .037$) and orientations (45° with 45° and 135° with 135°; mean $r = .0266$, $t(23) = 1.94$, $p = .065$). There was no reliable correlation for stimuli that were different, either for shapes (Shape 2 with Shape 4; mean $r = -.0041$, $t(23) = -0.31$, $p = .76$) or orientations (45° with 135°; mean $r = .0172$, $t(23) = 1.31$, $p = .20$). These results do not support the notion that the dissociation we report was caused by the two orientations, but not the shapes, evoking opposing patterns.

Can the difference between the shape and orientation experiments be explained by a difference in behavioral task performance/difficulty? Following the same logic as above, we split the participants in both experiments into two subgroups based on their behavioral accuracy. Specifically, we performed a median split on percentage correct responses, yielding high ($n = 12$) and low ($n = 12$) performers for both experiments. We investigated whether evidence for the predicted shape/orientation in the hippocampus differed between high versus low performers using a two-way ANOVA. This analysis revealed a main effect of experiment, $F(1, 44) = 9.00$, $p = .0044$; no main effect of task performance, $F(1, 44) = 0.03$, $p = .87$; and no interaction between the two, $F(1, 44) = 0.87$, $p = .36$. This fails to provide evidence that our main content-sensitive hippocampal effects can be attributed to differences in task difficulty.

Can the difference between the shape and orientation experiments be explained by a difference in patterns of eye movements? Another potential concern could be that the differential hippocampal effects could reflect predictive cues inducing different eye movements for shapes and gratings. In both experiments, participants were instructed to fixate on the bull's-eye in the center of the screen throughout the experiment and to not move their eyes toward the stimuli. Still, to examine potential influences of involuntary eye movements, we collected high-quality eye-tracking data for 9 of 24 participants in Experiment 1 and 15 of 24 in Experiment 2. We investigated influences of the stimulus (i.e., Shape 2 vs. Shape 4 in Experiment 1; 45° vs. 135° grating in Experiment 2), predicted stimulus, and the interaction of the two, on pupil position both poststimulus (250–750 msec) and during the cue–stimulus interval (–250 to 0 msec).

We found no evidence of the presented shape on pupil position in Experiment 1, either prestimulus (x-coordinate: $t(8) = -0.14$, $p = .89$; y-coordinate: $t(8) = 0.73$, $p = .49$) or poststimulus (x-coordinate: $t(8) = -1.44$, $p = .19$; y-coordinate: $t(8) = -1.10$, $p = .30$). In Experiment 2, there was a small difference in horizontal pupil position between 45° and 135° gratings poststimulus (mean x-coordinate = 0.01° vs. −0.05°, respectively; $t(14) = 2.21$, $p = .044$). This could reflect small involuntary eye movements along the orientation axes of the gratings (Mostert et al., 2018). There were no effects on prestimulus pupil position (x-coordinate: $t(14) = 1.95$, $p = .071$; y-coordinate: $t(14) = -1.32$, $p = .21$) or vertical poststimulus pupil position (y-coordinate: $t(14) = -1.21$, $p = .25$). However, crucial for the interpretation of our results is whether eye movements were influenced by the predictive cues. There were no effects of predicted shape, nor an interaction between predicted and presented shape, on pupil position in either pre- or poststimulus intervals (all ps $> .10$, both for horizontal and vertical pupil coordinates). Similarly, there were no effects of predicted orientation, nor an interaction between predicted and presented orientation, on pupil position in either pre- or poststimulus intervals (all ps $> .10$, both for horizontal and vertical pupil coordinates). That is, we found no evidence for differences in eye movements that could explain the fMRI effects of the predictive cues. In addition to specifying pre- and poststimulus time windows, we also conducted exploratory cluster-based permutation tests (Maris & Oostenveld, 2007) on the full-time window (−850 to 1000 msec). This analysis did not reveal any significant effects of presented or predicted stimulus, nor their interaction, for either Experiment 1 or 2 (no clusters $p < .05$).

It should be noted that these control analyses relied on splitting the participants into subgroups ($n = 12$ per group) or, in the case of the eye movement analysis, could only be performed on a subset of the participants ($n = 9$ and $n = 15$ for Experiments 1 and 2, respectively). Therefore, we cannot rule out the possibility that these analyses did not have sufficient power to detect some confounding effects.

DISCUSSION

Recent theories of the hippocampus suggest that it performs general-purpose computations independent of

stimulus contents (Buzsáki & Tingley, 2018). Alternatively, it has been suggested that the nature of stimuli, especially their complexity, is a crucial factor in determining whether hippocampus and MTL are involved in a given perceptual task (Dalton, Zeidman, McCormick, & Maguire, 2018; Murray et al., 2007). The current study addresses these hypotheses by revealing that predictions about complex shapes and simple grating orientations evoked qualitatively different representations in the hippocampus. This suggests that the hippocampus can play distinct computational roles in perception depending on the content of perceptual predictions, rather than executing a general-purpose process independent of stimulus content. This finding is especially noteworthy given that the experimental paradigms, fMRI scan sequences, and neural decoding methods were virtually identical in the two experiments. Furthermore, the effects of the predictions on processing in visual cortex were highly similar for complex shapes and oriented gratings, suggesting that the hippocampal differences were not due to simple differences in the extent to which the predictions were learned or used to guide perception.

Could our results have been caused by something other than the nature of the stimuli per se? First, shape and orientation predictions were measured in separate experiments, involving different participants and MR scanners. These factors were matched as well as possible by recruiting both participant populations from similar university campuses and by using the same type of MR scanner in both experiments. Second, the tasks were necessarily different in the two studies: detecting subtle shape warps (Experiment 1) versus subtle grating phase shifts (Experiment 2). These tasks were designed to be as similar as possible: Both involved detecting a subtle change in a feature that was orthogonal to the predicted feature. Future research is needed to address this limitation, for instance, by having participants perform the same distracting task at fixation across stimulus types. Third, could the dissociation be caused by the fact that orientation is not represented in hippocampus, whereas complex shapes are? A control analysis revealed no significant difference in decoding performance for shapes and orientation in hippocampus in the stimulus-only runs. Moreover, we report significant decoding of invalidly predicted orientations in CA2-CA3-DG, demonstrating that orientations are in fact decodable in hippocampus. Fourth, the shapes were sampled from a linear, noncircular space, whereas orientation is a circular feature space. Although control analyses suggest that this did not cause a qualitative difference in the patterns evoked in hippocampus by the two types of stimuli (e.g., opposing patterns for orientations, but orthogonal patterns for shapes), future work should address whether the circularity of the feature space affects the hippocampus's role in prediction. Finally, we performed several control analyses to investigate potential differences between the experiments in task difficulty, stimulus discriminability, or eye movements, none of which was able to explain our results.

In summary, the differential responses of the hippocampus in the two experiments seem best explained by the difference in the nature of the predicted stimuli: complex objects versus simple features. This is in line with theories on hierarchical message passing in the brain (Friston, 2005; Lee & Mumford, 2003; Rao & Ballard, 1999). Complex objects are known to be represented in the MTL, such as in perirhinal cortex (Martin et al., 2018; Murray et al., 2007), areas known to have direct reciprocal connections with the hippocampus (Henke, 2010; Lavenex & Amaral, 2000). Therefore, the hippocampus is ideally positioned to supply complex shape predictions to its immediate cortical neighbors. Specifically, upon reception of a predictive cue (rising or falling tones), pattern completion mechanisms in hippocampus, especially in the CA3 subfield (Schapiro et al., 2017; Hindy et al., 2016; Treves & Rolls, 1994), may lead to retrieval of the predicted associate, which can then be sent back to MTL cortex as a prediction of upcoming inputs. In contrast, for the low-level feature of orientation, hippocampus does not seem to represent the predicted feature. This may be explained by the fact that the nearby cortical recipients of hippocampal feedback in the MTL do not preferentially represent such low-level features. Rather, as reviewed above, these areas represent complex objects abstracted away from their simple features (i.e., invariant over location, size, etc.) and are thus not ideal targets for predictions about such features.

In fact, for oriented gratings, the hippocampus and especially its CA3 subfield (combined with CA2 and dentate gyrus) seemed to represent prediction "errors" rather than predictions (Duncan, Ketz, Inati, & Davachi, 2012; Chen, Olsen, Preston, Glover, & Wagner, 2011; Lisman & Grace, 2005): Validly cued orientations were cancelled out, whereas invalidly cued orientations were not. In other words, the hippocampus seemed to represent an "antiprediction" that inhibited representation of expected stimuli. This is consistent with the observed negative evidence for expected but omitted orientations. Such coding for stimulus prediction errors may allow the hippocampus to refine learning and predictions elsewhere in the brain. Note that we observed these effects in CA2-CA3-DG, whereas most theories propose that prediction errors or at least a comparison between retrieved and experienced information should occur in CA1 (Chen et al., 2011; Lisman & Grace, 2005). It is possible that we did not observe such CA1 effects because we scanned after associative learning of the cues and outcomes was complete and that they may be more apparent if we had examined responses during the learning process, a possibility that awaits future studies.

Note that, in the current experimental design, when one stimulus is predicted (e.g., Shape 2 in Experiment 1 or 45° grating in Experiment 2), the other (Shape 4 or 135° grating, respectively) is always the unpredicted

stimulus. Therefore, positive (negative) decoding evidence for the predicted stimulus could also reflect negative (positive) evidence for the unpredicted. Future research will be able to distinguish these two possibilities by increasing the number of cues and possible stimuli, such that the predicted and unpredicted stimuli can be dissociated.

Interestingly, shape prediction signals were strongly present in the subiculum, but orientation signals were fully absent there. The subiculum is a major output hub of hippocampus back to MTL cortex (Roy et al., 2017; Lavenex & Amaral, 2000). Therefore, this pattern of results is in line with the suggestion that hippocampus may be a top–down source for high-level object predictions in visual cortex, but not for low-level feature predictions. This proposal leads to distinct hypotheses about the direction of signal flow through the hippocampal and MTL system during processing of complex shape and feature predictions, respectively. That is, shape predictions are proposed to flow from the hippocampus (CA3 through subiculum) back to cortex via entorhinal cortex (EC), whereas orientation predictions are not (and prediction errors may flow forward from EC to CA1/CA3). These hypotheses can be tested in future research using layer-specific fMRI of EC (Koster et al., 2018; Maass et al., 2014), because signals flowing into hippocampus arise from superficial layers, whereas signals flowing from hippocampus back to cortex arrive in the deep layers (Lavenex & Amaral, 2000). Additionally, this can be addressed using simultaneous electrophysiological measurements in hippocampus and cortex, for instance, in human epilepsy patients, which offer superior temporal resolution.

When oriented gratings were expected but omitted, the pattern of activity in V1 reflected the expected orientation, suggesting that such expectations can evoke a template of the predicted feature in sensory cortex, in line with previous findings (Kok et al., 2014, 2017). In contrast, this did not occur for expected but omitted shapes. Together with the differential hippocampal representations, these findings suggest that expectations about low-level features and higher-level objects may involve distinct neural mechanisms. In this context, it is interesting that valid (vs. invalid) orientation predictions slightly improved participants' phase discrimination performance, in line with previous findings of improved orientation discrimination for validly predicted gratings (Kok et al., 2017; Kok, Jehee, et al., 2012), whereas we did not find such an improvement for the shape discrimination task. It should be noted that behavioral benefits are expected to be minor (or absent) in the current study, because participant's task (discriminating the two stimuli on a given trial) was orthogonal to the cue (which predicted the identity of the first stimulus on a trial).

As revealed by a searchlight analysis, this dissociation is not restricted to visual cortex and hippocampus but also occurs in anterior occipital cortex, cerebellum, left IFG, and left middle temporal gyrus (see Table 1). The exact role of these other regions in generating predictions and prediction errors is unknown, though it is interesting to note that the anterior occipital cortex (St John-Saaltink, Utzerath, Kok, Lau, & De Lange, 2015), cerebellum (Roth, Synofzik, & Lindner, 2013), and left IFG (Turk-Browne, Scholl, Chun, & Johnson, 2009) have previously been implicated in perceptual prediction and statistical learning. Further work is required to establish the hierarchy of regions involved in generating expectations. For instance, the left IFG is a high-level region with the appropriate connectivity for sending top–down signals to sensory cortex, and it would be of great interest to know whether the negative orientation prediction signals there originate in the hippocampus, or whether instead the hippocampus receives these signals from left IFG.

The effects of the orientation predictions on grating-evoked signals in visual cortex, as reported here, differ from those reported previously using a similar paradigm (Kok, Jehee, 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, Jehee, et al., 2012), the current study found that invalid orientation predictions lead to "delayed" signals, both in terms of BOLD amplitude and orientation representations. Although the cause of this difference is currently unclear, there were a couple of potentially important differences between these studies. First, the two studies employed different behavioral tasks: Participants performed orientation and contrast discrimination in Kok, Jehee, et al. (2012), whereas in the current study they discriminated grating phase. Although we do not have a clear hypothesis for how this would lead to differences in V1, previous work has shown that task demands influence expectation effects in visual cortex (Auksztulewicz, Friston, & Nobre, 2017; St John-Saaltink et al., 2015; Kok, Rahnev, Jehee, Lau, & De Lange, 2012; Larsson & Smith, 2012). Additionally, there were differences in the grating presentations between the two studies. In Kok, Jehee, et al. (2012), the individual gratings were presented for a longer duration (500 vs. 250 msec) but with a shorter ISI (100 vs. 500 msec), and the two gratings in a given trial were in antiphase with different spatial frequencies. Whether and how these parameters could explain the differential effects of expectation cues on V1 processing is unclear, but one possibility is that they might affect the degree of repetition suppression between the two gratings in each trial, which could in turn interact with prediction signals (Henson, 2016; Kok, Jehee, et al., 2012; Todorovic & De Lange, 2012; Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008).

In summary, the current study revealed that predictions about complex shapes and simple orientations evoke distinct representations in hippocampus. These findings are in line with the hippocampus generating perceptual predictions for high-level objects, but not

for low-level features. This fits well with hierarchical Bayesian inference theories of sensory processing (Friston, 2005; Lee & Mumford, 2003; Rao & Ballard, 1999), which suggest that each brain region provides predictions to those regions with which it has direct feedback connections and formats those predictions in the currency that the receiving region "understands" (Bastos et al., 2012; Lee & Mumford, 2003). Finally, these findings suggest that stimulus complexity is a crucial factor in determining whether and in what role the hippocampus is involved in perceptual inference (Murray et al., 2007).

Acknowledgments

This work was supported by an NWO Rubicon grant 446-15-004 to P. K. and NIH R01 MH069456 to N. B. T.-B. The Wellcome Centre for Human Neuroimaging is supported by core funding from the Wellcome Trust (203147/Z/16/Z).

Reprint requests should be sent to Peter Kok, Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology, 12 Queen Square, London WC1N 3AR, United Kingdom, or via e-mail: p.kok@ucl.ac.uk.

REFERENCES

Aly, M., & Turk-Browne, N. B. (2016a). Attention stabilizes representations in the human hippocampus. *Cerebral Cortex*, *26*, 783–796.

Aly, M., & Turk-Browne, N. B. (2016b). Attention promotes episodic encoding by stabilizing hippocampal representations. *Proceedings of the National Academy of Sciences, U.S.A.*, *113*, E420–E429.

Andersson, J. L., 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.

Auksztulewicz, R., Friston, K. J., & Nobre, A. C. (2017). Task relevance modulates the behavioural and neural effects of sensory predictions. *PLoS Biology*, *15*, e2003143.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*, 695–711.

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., et al. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron*, *100*, 490–509.

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

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *Journal of Neuroscience*, *29*, 13992–14003.

Brouwer, G. J., & Heeger, D. J. (2011). Cross-orientation suppression in human visual cortex. *Journal of Neuroscience*, *106*, 2108–2119.

Buzsáki, G., & Tingley, D. (2018). Space and time: The hippocampus as a sequence generator. *Trends in Cognitive Sciences*, *22*, 853–869.

Chen, J., Olsen, R. K., Preston, A. R., Glover, G. H., & Wagner, A. D. (2011). Associative retrieval processes in the human medial temporal lobe: Hippocampal retrieval success and CA1 mismatch detection. *Learning & Memory*, *18*, 523–528.

Cowell, R. A., Barense, M. D., & Sadil, P. S. (2019). A roadmap for understanding memory: Decomposing cognitive processes into operations and representations. *eNeuro*, *6*, ENEURO.0122-19.2019.

Cowell, R. A., Leger, K. R., & Serences, J. T. (2017). Feature-coding transitions to conjunction-coding with progression through human visual cortex. *Journal of Neurophysiology*, *118*, 3194–3214.

Dalton, M. A., Zeidman, P., McCormick, C., & Maguire, E. A. (2018). Differentiable processing of objects, associations, and scenes within the hippocampus. *Journal of Neuroscience*, *38*, 8146–8159.

Davachi, L., & DuBrow, S. (2015). How the hippocampus preserves order: The role of prediction and context. *Trends in Cognitive Sciences*, *19*, 92–99.

De Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*, 764–779.

Drucker, D. M., & Aguirre, G. K. (2009). Different spatial scales of shape similarity representation in lateral and ventral LOC. *Cerebral Cortex*, *19*, 2269–2280.

Duncan, K., Ketz, N., Inati, S. J., & 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.

Eichenbaum, H., & Fortin, N. J. (2009). The neurobiology of memory based predictions. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *364*, 1183–1191.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.

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

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-related fMRI: Characterizing differential responses. *Neuroimage*, *7*, 30–40.

Garvert, M. M., Dolan, R. J., & Behrens, T. E. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *eLife*, *6*, e17086.

Hafting, T., Fyhn, M., Molden, S., Moser, M. B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*, 801–806.

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience*, *11*, 523–532.

Henson, R. N. (2016). Repetition suppression to faces in the fusiform face area: A personal and dynamic journey. *Cortex*, *80*, 174–184.

Henson, R. N., Price, C. J., Rugg, M. D., Turner, R., & Friston, K. J. (2002). Detecting latency differences in event-related BOLD responses: Application to words versus nonwords and initial versus repeated face presentations. *Neuroimage*, *15*, 83–97.

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.

Hindy, N. C., & Turk-Browne, N. B. (2016). Action-based learning of multistate objects in the medial temporal lobe. *Cerebral Cortex*, *26*, 1853–1865.

Jabar, S. B., Filipowicz, A., & Anderson, B. (2017). Tuned by experience: How orientation probability modulates early perceptual processing. *Vision Research*, *138*, 86–96.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). FSL. *Neuroimage*, *62*, 782–790.

Kaposvari, P., Kumar, S., & Vogels, R. (2018). Statistical learning signals in macaque inferior temporal cortex. *Cerebral Cortex*, *28*, 250–266.

Kok, P., Brouwer, G. J., Van Gerven, M. A., & De Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *Journal of Neuroscience*, *33*, 16275–16284.

Kok, P., Failing, M. F., & De Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of Cognitive Neuroscience*, *26*, 1546–1554.
Kok, P., Jehee, J. F., & De Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron*, *75*, 265–270.
Kok, P., Mostert, P., & De Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences, U.S.A.*, *114*, 10473–10478.
Kok, P., Rahnev, D., Jehee, J. F., Lau, H. C., & De Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex*, *22*, 2197–2206.
Kok, P., & Turk-Browne, N. B. (2018). Associative prediction of visual shape in the hippocampus. *Journal of Neuroscience*, *38*, 6888–6899.
Koster, R., Chadwick, M. J., Chen, Y., Berron, D., Banino, A., Düzel, E., et al. (2018). Big-loop recurrence within the hippocampal system supports integration of information across episodes. *Neuron*, *99*, 1342–1354.
Larsson, J., & Smith, A. T. (2012). fMRI repetition suppression: Neuronal adaptation or stimulus expectation? *Cerebral Cortex*, *22*, 567–576.
Lavenex, P., & Amaral, D. G. (2000). Hippocampal-neocortical interaction: A hierarchy of associativity. *Hippocampus*, *10*, 420–430.
Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, *20*, 1434–1448.
Lisman, J., & Redish, A. D. (2009). Prediction, sequences and the hippocampus. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *364*, 1193–1201.
Lisman, J. E., & Grace, A. A. (2005). The hippocampal–VTA loop: Controlling the entry of information into long-term memory. *Neuron*, *46*, 703–713.
Liu, K., Sibille, J., & Dragoi, G. (2018). Generative predictive codes by multiplexed hippocampal neuronal tuplets. *Neuron*, *99*, 1329–1341.
Maass, A., Schütze, H., Speck, O., Yonelinas, A., Tempelmann, C., Heinze, H. J., et al. (2014). Laminar activity in the hippocampus and entorhinal cortex related to novelty and episodic encoding. *Nature Communications*, *5*, 5547.
Manahova, M. E., Mostert, P., Kok, P., Schoffelen, J. M., & De Lange, F. P. (2018). Stimulus familiarity and expectation jointly modulate neural activity in the visual ventral stream. *Journal of Cognitive Neuroscience*, *30*, 1366–1377.
Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190.
Martin, C. B., Douglas, D., Newsome, R. N., Man, L. L., & Barense, M. D. (2018). Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. *eLife*, *7*, e31873.
McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (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. *Psychological Review*, *102*, 419–457.
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.
Mostert, P., Albers, A. M., Brinkman, L., Todorova, L., Kok, P., & De Lange, F. P. (2018). Eye movement-related confounds in neural decoding of visual working memory representations. *eNeuro*, *5*, ENEURO.0401-17.2018.
Murray, E. A., Bussey, T. J., & Saksida, L. M. (2007). Visual perception and memory: A new view of medial temporal lobe function in primates and rodents. *Annual Review of Neuroscience*, *30*, 99–122.
Op de Beeck, H., Wagemans, J., & Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nature Neuroscience*, *4*, 1244–1252.
Rao, R. P., & 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.
Recasens, M., Gross, J., & Uhlhaas, P. J. (2018). Low-frequency oscillatory correlates of auditory predictive processing in cortical–subcortical networks: A MEG-study. *Scientific Reports*, *8*, 14007.
Richter, D., Ekman, M., & De Lange, F. P. (2018). Suppressed sensory response to predictable object stimuli throughout the ventral visual stream. *Journal of Neuroscience*, *38*, 7452–7461.
Roth, M. J., Synofzik, M., & Lindner, A. (2013). The cerebellum optimizes perceptual predictions about external sensory events. *Current Biology*, *23*, 930–935.
Roy, D. S., Kitamura, T., Okuyama, T., Ogawa, S. K., Sun, C., Obata, Y., et al. (2017). Distinct neural circuits for the formation and retrieval of episodic memories. *Cell*, *170*, 1000–1012.
Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*, *22*, 1622–1627.
Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *372*, 20160049.
Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, *44*, 83–98.
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.
Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, *20*, 1643–1653.
Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*, 1004–1006.
Todorovic, A., & De Lange, F. P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *Journal of Neuroscience*, *32*, 13389–13395.
Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*, 374–391.
Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J. J., Moser, M.-B., et al. (2018). Integrating time from experience in the lateral entorhinal cortex. *Nature*, *561*, 57–62.
Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*, 1934–1945.
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.

Utzerath, C., St. John-Saaltink, E., Buitelaar, J., & De Lange, F. P. (2017). Repetition suppression to objects is modulated by stimulus-specific expectations. *Scientific Reports*, *7*, 8781.

Wang, R., Shen, Y., Tino, P., Welchman, A. E., & Kourtzi, Z. (2017). Learning predictive statistics: Strategies and brain mechanisms. *Journal of Neuroscience*, *37*, 8412–8427.

Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*, 113–120.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *Neuroimage*, *92*, 381–397.

Yonelinas, A. P. (2013). The hippocampus supports high-resolution binding in the service of perception, working memory and long-term memory. *Behavioural Brain Research*, *254*, 34–44.

Yushkevich, P. A., Pluta, J. B., Wang, H., Xie, L., Ding, S.-L., Gertje, E. C., et al. (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. *Human Brain Mapping*, *36*, 258–287.

Zahn, C. T., & Roskies, R. Z. (1972). Fourier descriptors for plane closed curves. *IEEE Transactions on Computers*, *C-21*, 269–281.