Published in final edited form as:

Nat Neurosci. 2016 May ; 19(5): 665–667. doi:10.1038/nn.4284.

Linking pattern completion in the hippocampus to predictive coding in visual cortex

Nicholas C. Hindy[1,*], **Felicia Y. Ng**[2], and **Nicholas B. Turk-Browne**[1,2]

[1]Princeton Neuroscience Institute, Princeton University

[2]Department of Psychology, Princeton University

Abstract

Models of predictive coding frame perception as a generative process in which expectations constrain sensory representations. These models account for expectations about how a stimulus will move or change from moment to moment, but do not address expectations about what other, distinct stimuli are likely to appear based on prior experience. We show that such memory-based expectations in human visual cortex are related to the hippocampal mechanism of pattern completion.

At least two kinds of expectations guide perception. First, we form 'perceptual' expectations about how current stimuli move or change over time. For example, when driving, we anticipate the locations of signs and cars in our field-of-view at the next moment. Hierarchical models of predictive coding explain how such expectations arise, with feedback signals carrying information about expected levels of activity in earlier layers of visual cortex[1–3]. These signals modulate sensory representations, accounting for neurophysiological findings in areas V1 and V2 such as contour filling-in[4] and motion anticipation[5].

Second, we form 'mnemonic' expectations about what new stimuli are likely to appear in the near future. For example, when turning at a familiar intersection, we anticipate the identities of buildings and streets that will come into view. Mnemonic expectations differ from perceptual expectations because expected stimuli need not physically resemble current stimuli and are instead expected based on prior co-occurrence[6,7]. In the example above, the name of the street you turned on has no inherent connection to the look of an upcoming building, despite the former being predictive of the latter on a known route. Mnemonic

*Corresponding author: Nicholas Hindy, Peretsman-Scully Hall 324, Princeton, NJ, 08544, nhindy@princeton.edu.

Author Contributions

N. Hindy, F. Ng, and N. Turk-Browne designed the experiment, reviewed the analyses, and discussed the results. N. Hindy and F. Ng collected the data. N. Hindy performed the analyses. N. Hindy and N. Turk-Browne wrote and revised the paper.

Competing Financial Interests

The authors declare no competing financial interests.

Code availability

Data and code are available upon request from the first author (nhindy@princeton.edu).

expectations can influence activity in the same areas of early visual cortex as perceptual expectations[8].

Predictive coding models say little about how mnemonic expectations arise. The often-arbitrary nature of mnemonic expectations requires a different mechanism than feedback from adjacent areas. Specifically, mnemonic expectations require retrieval of past experiences in order to anticipate upcoming information in sensory areas. One candidate retrieval mechanism is pattern completion in the hippocampus[9–11], whereby exposure to part of a past experience activates a conjunctive representation of the entire experience. This occurs through recurrent connectivity in the CA3 subfield of the hippocampus, which allows activity to spread from the partial input to other inputs with which it was bound during encoding. The completed representation is transferred to the CA1 subfield and then is output to cortical regions, with visual components reinstated in areas V1 and V2 of early visual cortex[12,13]. As an initial step toward establishing a role for pattern completion in predictive coding, we test the relationship between representations in CA3–CA1 and V1–V2 during mnemonic expectation.

We first trained human participants to expect a specific outcome stimulus when they performed a particular motor action in response to a visual cue (Supplementary Fig. 1)[14]. After training, we used high-resolution functional MRI (fMRI) to measure patterns of activity elicited by trials in which a cue was presented and an action was performed, but no outcome was received (Fig. 1a). We hypothesized that such cue + action trials would trigger both pattern completion and predictive coding, with the hippocampus more involved in pattern completion and visual cortex more involved in predictive coding. That is, upon acting on the cue, CA3 and CA1 may retrieve a conjunctive representation of the corresponding full cue + action + outcome sequence, in turn setting up an expectation of the outcome by reinstating it in V1 and V2.

To operationalize pattern completion on cue + action trials, we trained a multivariate classifier on patterns of blood oxygenation level–dependent contrast (BOLD) activity from full-sequence trials (sequence decoding), as these trials contained the most retrieval cues for eliciting conjunctive representations of the sequences (Fig. 1b). To operationalize predictive coding on cue + action trials, we trained another classifier on outcome-only trials (outcome decoding), as these trials provided the purest assay of the stimulus representations of the outcomes (Fig. 1c). Separate classifiers were trained in CA3 (including CA2 and dentate gyrus (DG)), CA1, V1, and V2 regions of interest (ROIs; Fig. 2a).

According to our hypothesis, the hippocampus should be more likely to show sequence decoding and visual cortex should be more likely to show outcome decoding. Indeed, there was an interaction between classifier type and region ($F_{1,23} = 8.97$, $P = 0.006$). Sequence decoding was reliable in CA2–CA3–DG and CA1, but not in V1 or V2 (Fig. 2b, Supplementary Fig. 2a). Outcome decoding was reliable in V1 and V2, but not in CA2–CA3–DG or CA1 (Fig. 2c, Supplementary Fig. 2b). Both sequence decoding in the hippocampus and outcome decoding in visual cortex were eliminated in a control condition that closely matched the experimental condition except that the actions were no longer predictive of outcomes (Supplementary Fig. 3). Moreover, several control analyses of

predictive and non-predictive actions were inconsistent with the possibility that actions per se were sufficient for sequence decoding in the hippocampus (Supplementary Figs. 4 and 5). Finally, the strength of each effect was related across participants to performance on behavioral tests of learning outside of the scanner (Supplementary Fig. 6).

The complete lack of sequence decoding in visual cortex and outcome decoding in the hippocampus for predictive actions is notable. Indeed, full-sequence trials contained different visual outcomes, which might have enabled sequence decoding in visual cortex, and outcome-only trials provided some basis for pattern completion, which might have enabled outcome decoding in the hippocampus. A post hoc cross-classification analysis, which compared all combinations of trial types as training and testing data, supported the interpretation that classification was successful when there was a match between the type of information represented in a region and the type of information most discriminative across classifier examples (Supplementary Fig. 7). Because visual cortex represents visual stimuli, a classifier may have trouble discriminating trials with a common stimulus (i.e., the shared cue on full-sequence trials), as this stimulus could induce counter-productive neural similarity. Because the hippocampus represents conjunctions, a classifier may have trouble discriminating trials with only one element (i.e., the outcome on outcome-only trials), as this element could lead to weak retrieval and noisy neural patterns.

Having found sequence decoding in the hippocampus and outcome decoding in visual cortex, we next asked about the relationship between these effects. Consistent with our hypothesis that pattern completion in the hippocampus may underlie predictive coding in visual cortex, outcome decoding in a combined V1–V2 ROI was more accurate within participants on cue + action trials in which the correct vs. incorrect sequence was decoded in a combined hippocampal (CA–DG) ROI (Fig. 3a, Supplementary Fig. 8a). This relationship also held across participants, with average outcome decoding in V1–V2 positively correlated with average sequence decoding in CA–DG (Fig. 3b, Supplementary Fig. 8b). Both effects were eliminated when the analyses were reversed to relate sequence decoding in V1–V2 to outcome decoding in CA–DG ($P > 0.39$).

These correlational links between the hippocampus and visual cortex do not address the directionality of the relationship. This cannot be established definitively with fMRI—invasive studies would be needed—but as a suggestive first step we examined the relative timing of sequence information in the hippocampus and outcome information in visual cortex. Using multinomial regression, we found that sequence decoding in the hippocampus early in a cue + action trial was predictive of outcome decoding in visual cortex later in that trial (Supplementary Fig. 9). This cross-correlation was asymmetric in time, as the reverse correlation (visual cortex predicting hippocampus) was not reliable. Such dynamics suggest that sequence information in the hippocampus preceded outcome information in visual cortex, consistent with the hippocampus reinstating expected outcomes in visual cortex.

Overall, our findings suggest that hippocampal pattern completion may provide a mechanism for action-based mnemonic expectation and predictive coding more generally[8,13]. Although the hippocampus is one potential—and theoretically grounded—source of predictions, the specificity of its contribution to predictive coding remains an open

question. We only obtained partial coverage of the brain, and other systems have repeatedly been linked to prediction, including the ventral striatum[15] and orbitofrontal cortex[16].

Online Methods

Participants

Twenty-four individuals (13 females and 11 males, aged 18–30 years) participated in the study. The effect size of interest was not known in advance, so sample size was chosen to match a previous fMRI study with a similar behavioral design[14]. Each participant was right-handed and had normal or corrected-to-normal vision. Two additional participants were removed from the scanner before completing the experiment (because of fatigue and excessive movement, respectively), and were excluded from data analysis. Participants were recruited from the Princeton University community and were paid $20 per hour. Informed consent was obtained using a protocol approved by the Princeton University Institutional Review Board.

Stimuli

The stimulus set of 12 fractal-like images is displayed in Supplementary Figure 1. The images were created using ArtMatic Pro (http://www.artmatic.com), and subtended approximately 4 degrees of visual angle in diameter on the laptop computer used for behavioral training and testing and 4.5 degrees in the scanner. Images were randomly assigned to serve as cues or outcomes.

Behavioral training

Training consisted of two 40-min sessions. The first session occurred approximately 24 h before scanning, and the second session occurred immediately before scanning. The first session and the first half of the second session were performed on a laptop computer. The second half of the second session was performed in the scanner during structural imaging, to familiarize participants with the appearance of stimuli in this new environment. Each training session included 336 full-sequence trials, with 84 trials for each of two predictable cues and 84 trials for each of two unpredictable cues. Participants were instructed to discover which of the two possible outcomes for each cue was most likely to appear after a button press with the index finger of the left hand, and which was most likely to appear after a button press with the right index finger. To acquaint participants with the trial types that would later appear during the scan task, each training session additionally included four cue + action trials and eight outcome-only trials.

For each full-sequence trial of the first ‘exploratory’ training session, participants were shown a cue stimulus for 1,000 ms and then a double-headed arrow appeared below the cue. This prompted them to decide which action to perform. Upon pressing a button with their left or right hand, the cue stimulus was replaced by an outcome stimulus. A meter at the bottom of the screen tracked the proportion of left and right button presses during the first training session, and participants were instructed to keep the meter within a specified central zone, in order to roughly equate the frequency of actions and outcomes. During the second ‘directed’ training session, a single-headed arrow was shown after the onset of the cue,

which instructed participants to perform the left or right action. This was done so we could equate the stimulus frequencies and transition probabilities of the two outcomes associated with each cue throughout training. For example, if participants responded left more than right during the exploratory training, they were more likely to be instructed to respond right in the directed training.

For each participant, four different cue stimuli were each associated with two unique outcome stimuli. Two of these cue–outcome stimulus triads were assigned to the predictable condition: given cue A, outcome B appeared with 95% probability when the left button was pressed and outcome C appeared with 95% probability when the right button was pressed; on the remaining 5% of trials, the outcomes were swapped. The other two cue–outcome stimulus triads were assigned to the unpredictable condition: given cue D, outcomes E and F each appeared with 50% probability when either the left or right button was pressed. Thus, actions were meaningless for unpredictable triads, as they did not provide any information about which outcome would appear. Since unpredictable trials were otherwise identical to predictable trials, they served as a baseline control for task components unrelated to action-based prediction, such as button presses and the learning of stimulus-stimulus associations.

Behavioral tests

To verify learning of the full cue + action + outcome sequences, each participant performed two pre-scan behavioral tests and one post-scan behavioral test. On each test trial, a cue stimulus appeared at fixation. Below the cue, a single-headed arrow pointed left or right, instructing participants to press the corresponding button. The cue and arrow then disappeared, replaced by the two possible outcomes for that cue, presented above and below where the cue had been. One outcome correctly completed the sequence given the performed action, while the other outcome completed the sequence for the other non-performed action. Participants had a 4 s response window to indicate which outcome was expected by saying aloud either “top’ or “bottom’. Verbal response was used to avoid the button response actions that were an important part of the training. In a pre-scan test that followed the first training session, participants were required to achieve 100% accuracy in the predictable condition, or they repeated the training until they reached perfection. Accuracy for predictable sequences was 99.0% on average (s.d. = 3.5%) in a pre-scan test that followed the second training session, and 98.4% on average (s.d. = 5.6%) in the post-scan test. Both means were robustly above the chance level of 50% ($P < 0.001$).

Audio signal for the behavioral tests was sampled at 44.1 kHz, and voice response time (RT) for each predictable and unpredictable sequence was measured as the timestamp of the first audio sample in which the absolute value of the signal amplitude was greater than 50% of the maximum amplitude within the 4 s response window. Average RT was lower overall ($t_{23} = 2.98$, $P = 0.007$) for predictable trials (mean = 1,088 ms, s.d. = 248 ms) than for unpredictable trials (mean = 1,241 ms, s.d. = 382 ms).

Scan task

Eight fMRI runs lasting about six minutes each were collected. A total of 576 trials were equally distributed across the runs. This total breaks down into three randomly intermixed

trial types: 256 full-sequence trials, 128 cue + action trials, and 192 outcome-only trials. The trials of each type were evenly split between predictable and unpredictable conditions. The full-sequence trials resembled the first training session: a cue stimulus for 1,000 ms, followed by a double-headed arrow below the cue for up to 1,500 ms that prompted participants to choose and perform an action, and then an outcome stimulus for 1,000 ms immediately after the button press. The cue + action trials were identical to the full-sequence trials, except that the outcome stimulus was replaced with a blank screen for 1,000 ms after the button press. The outcome-only trials contained just the outcome stimulus for 1,000 ms, without a preceding cue or action. Participants used a separate response box for each hand to make the left or right button presses. If a button was not pressed in the 1,500 ms response window, the cue stimulus and action prompt were replaced with a fixation cross that remained on screen until the next trial. The order of trial types and the interstimulus intervals (ISIs) in each run were optimized for statistical power using optseq[2] (https://surfer.nmr.mgh.harvard.edu/optseq)[17]. The average ISI was 3,612 ms, which included a fixation interval of 1,500, 3,000, or 4,500 ms, plus the remaining time from the response window in the previous trial (1,500 ms minus the RT).

MRI acquisition

Structural and functional MRI data were collected on a 3T Siemens Skyra scanner with a 16-channel head coil. Structural data included a T1-weighted magnetization prepared rapid acquisition gradient-echo (MPRAGE) sequence (1 mm isotropic) for registration and segmentation of early visual cortex, and two T2-weighted turbo spin-echo (TSE) sequences (0.44 × 0.44 × 1.5 mm) for hippocampal segmentation. Functional data consisted of T2*-weighted multi-band echo-planar imaging sequences with 42 oblique slices (16° transverse to coronal) acquired in an interleaved order (1,500 ms repetition time (TR), 40 ms echo time, 1.5 mm isotropic voxels, 128 × 128 matrix, 192 mm field of view, 71° flip angle, acceleration factor 3, shift 2). These slices produced only a partial volume for each participant, parallel to the hippocampus and covering the temporal and occipital lobes. Collecting a partial volume instead of the full brain allowed us to maximize spatial and temporal resolution over our a priori ROIs. However, this prevented us from evaluating the selectivity of the findings elsewhere in the brain. Data acquisition in each functional run began with 12 s of rest in order to approach steady-state magnetization. A B0 field map was collected at the end of the experiment.

Regions of interest

Hippocampal subfields, including CA2–CA3–DG, CA1, and the subiculum, were defined in the TSE images using the automatic segmentation of hippocampal subfields (ASHS) machine learning toolbox[18] and a database of manual medial temporal lobe (MTL) segmentations from a separate set of 24 participants[19]. Manual segmentations were based on anatomical landmarks used in prior studies[6,19–21]. Consistent with these studies, CA2, CA3 and DG were combined into a single ROI because these subfields cannot be distinguished at our functional resolution (1.5 mm isotropic). The inclusion of DG could in principle be problematic for observing pattern completion because it is often linked to the opposite effect of pattern separation[22]. However, DG may also support pattern completion, as suggested by recent computational models[23] and neurophysiological findings[24]. We had reason to believe

that CA1 might show pattern completion, as it receives input from CA3 via Schaffer collaterals, and is believed to translate completed representations such that they can be reactivated in cortex[11]. Indeed, functional connectivity between CA3 and CA1 is enhanced during retrieval[25] and evidence of pattern completion has been observed in CA1. Not much is known about the role of the subiculum in pattern completion and this region is left out of most hippocampal models[9] and theories[11]. We included it as a control region where pattern completion might not be observed, as well as for the sake of completeness and to mirror prior high-resolution studies of human hippocampus[6,19,21,25,26]. Finally, in visual cortex, we focused on V1 and V2 because they can be precisely segmented anatomically within individual participants[27,28]. These ROIs were automatically defined in each participant's T1-weighted anatomical scan with FreeSurfer (http://surfer.nmr.mgh.harvard.edu/)[29].

We used two approaches to examine pattern completion and predictive coding elsewhere in our field of view. First, we defined V3 and V4 using a different probabilistic atlas[30], though this was less precise than for V1–V2 because it was done in Montreal Neurological Institute (MNI) space, to which each participant was registered. We did not obtain reliable outcome decoding in either V3 (t_{23} = 1.48, P = 0.15) or V4 (t_{23} = 1.23, P = 0.23), nor reliable sequence decoding in V3 (t_{23} = 0.30, P = 0.77) or V4 (t_{23} = 0.12, P = 0.91). Second, we centered a spherical multivariate searchlight31 with a 3-voxel (4.5 mm) radius on every voxel. In each searchlight, we trained and tested full-sequence and outcome-only classifiers as in the ROIs. We compared the resulting classifier accuracies to 50% chance using t-tests across participants, and assigned the resulting statistic to the center voxel. Group searchlight maps were corrected for multiple comparisons at $P < 0.05$, with a cluster-forming voxelwise α of $P < 0.001$ and a cluster-size threshold from 3dClustSim (http://afni.nimh.nih.gov/pub/dist/doc/program_help/3dClustSim.html) based on the smoothness of each map from 3dFWHMx (http://afni.nimh.nih.gov/pub/dist/doc/program_help/3dFWHMx.html). The sequence decoding searchlight analysis produced one reliable cluster in right putamen ($P < 0.05$ corrected; center-of-mass MNI coordinates: x = 34, y = −11, z = 0). The outcome decoding searchlight analysis did not produce any reliable clusters. These findings suggest that the ROI analyses did not obscure smaller patches of sequence information in V1–V2 or outcome information in the hippocampus, and that we benefitted from the greater statistical power and precision afforded by using a small number of a priori ROIs.

We further interrogated the right putamen in a series of control analyses to better understand the nature of sequence decoding in the hippocampus. Namely, because the right putamen has previously been linked to motor sequence learning[32,33], we interpreted its ability to discriminate between sequences as resulting from the different left/right motor actions in each sequence pair rather than from pattern completion of different conjunctive representations of the cue, action and outcome. In order to obtain a region of more comparable size to the hippocampal ROIs, we defined the right putamen anatomically from the Harvard-Oxford Subcortical Atlas.

Preprocessing

Data were preprocessed and spatially registered using the Oxford Centre for Functional MRI of the Brain (FMRIB) Software Library 5.0 (FSL5)[34]. Functional runs were corrected for

slice-acquisition time and head motion, high-pass filtered in time using a 50 s period cutoff, and spatially smoothed using a 3 mm full-width half-maximum (FWHM) Gaussian kernel. These runs were also registered to each participant's MPRAGE image using boundary-based registration[35] with B0-fieldmap correction, and then through FMRIB's Linear Image Registration Tool (FLIRT)[36] to the TSE images used for anatomical segmentation of hippocampal subfields. Primary ROI analyses were performed in each participant's native space. For other analyses, functional runs and high-resolution MPRAGE images were registered through FMRIB's Non-linear Image Registration Tool (FNIRT)[37] to the MNI152 template (Montreal Neurological Institute), which had been resampled via interpolation to match the resolution of the functional data (1.5 mm isotropic).

GLM

General linear model. Beta coefficients reflecting BOLD responses during the scan task were estimated with a general linear model (GLM) in FMRIB's Improved Linear Model (FILM)[34], which included temporal autocorrelation correction and six motion parameters as nuisance covariates. Each trial was modeled individually with a boxcar that lasted 1,000 ms for outcome-only trials and that matched the participant's average trial duration for full-sequence and cue + action trials (between 2,500 and 2,600 ms, depending on RT), and then convolved with a double-gamma hemodynamic response function. This resulted in a spatial map of parameter estimates for each trial in every condition that served as input to the classifiers. There was no difference in RT between predictable and unpredictable trials ($F_{1,23}$ = 0.69, P = 0.42) or between full-sequence and cue + action trials ($F_{1,23}$ = 0.34, P = 0.57).

MVPA

Multivariate pattern analysis. Beta coefficients were extracted from ROIs with MATLAB, and multivariate pattern analysis (MVPA) was performed using the Princeton MVPA Toolbox (http://www.pni.princeton.edu/mvpa). For each analysis, vectors of parameter estimates were z-scored within voxel across examples and then across voxels within each ROI or searchlight, and a logistic regression with L2-norm regularization (penalty = 1) was used as the classifier algorithm. Within each ROI or searchlight, and for each of the two predictable cues and each of the two unpredictable cues, one classifier was trained to distinguish the two alternative sequences using the full-sequence trials and a separate classifier was trained to distinguish the two alternative outcomes using the outcome-only trials. All classifiers were tested on the cue + action trials. The classification accuracies during testing for the two cues from each condition and classifier type were averaged to produce estimates of sequence decoding and outcome decoding, respectively. We used 24 training examples per outcome (48 total per outcome-only classifier); however, the number of training examples per sequence varied because participants chose which button to press for each trial (mean = 32 examples; minimum = 13 examples). The more and less frequent sequences for each cue always summed to 64 examples per full-sequence classifier. Note that the imbalanced training set across classes (and similarly imbalanced testing set for cue + action trials) was constant across ROIs, and therefore cannot account for the regional differences we observe.

Classification approach

Primary classification analyses were designed based on the match between the type of information most strongly represented in each ROI and the type of representation most strongly elicited by each trial type. Our choice of ROIs was motivated by prior evidence of conjunctive representations in the hippocampus[23,38,39] and top-down expectations in early visual cortex[8,13,40]. This led to two premises about how the information in our design would be represented in these regions. First, the hippocampus should form a conjunctive representation of each cue + action + outcome combination that is repeatedly experienced within a sequence. Second, early visual cortex should represent visual components of these sequences, namely the cue and outcome stimuli. The different trial types were included in order to elicit different neural representations. We reasoned that full-sequence trials would best elicit conjunctive representations because each trial was a direct repetition of a sequence and therefore provided the most retrieval cues (i.e., the cue, action and outcome). We reasoned that outcome-only trials would produce the purest neural representation of individual outcomes because each trial contained only the outcome stimulus—not other visual stimuli shared across sequences (i.e., the cue on full-sequence and cue + action trials).

Combining these logical steps resulted in two key hypotheses about where in the brain different classifiers would succeed in decoding neural patterns for cue + action trials. First, a classifier trained on patterns of activity in the hippocampus from the full-sequence trials would learn to distinguish the conjunctive representations of the two sequences for each cue. Insofar as cue + action trials elicit these conjunctive representations via pattern completion, this full-sequence classifier should be able to decode the identity of a sequence when tested on a cue + action trial (sequence decoding). Second, a classifier trained on patterns of activity in early visual cortex from the outcome-only trials would learn to distinguish the sensory representations of the two outcome stimuli for each cue. Insofar as cue + action trials elicit these outcome representations via predictive coding, this outcome-only classifier should be able to decode the identity of an outcome when tested on a cue + action trial (outcome decoding). In summary, the crossover interaction (hypothesized and observed) between sequence decoding in the hippocampus (Fig. 2b; Supplementary Fig. 2a) and outcome decoding in early visual cortex (Fig. 2c; Supplementary Fig. 2b) depend upon a combination of where in the brain different kinds of information were represented, the type of information elicited by different trials, and which trials are included as training and testing data for different classifiers.

Unpredictable trials

Analyses of the unpredictable trials addressed alternative accounts of the classification findings from the predictable trials, namely that they reflected decoding of left vs. right button presses or other task components. Note that sequence decoding and outcome decoding for predictable trials were based on the probability of a classifier guessing the correct full sequence or outcome, respectively, given a cue+action. For unpredictable trials, however, there is no objectively correct answer, as each cue+action was equally associated (by definition) with both full sequences and both outcomes. To obtain classification accuracy, we therefore relied on a subjectively “correct” answer defined from the behavioral tests performed outside the scanner. For each non-predictive cue, we identified the

participant's idiosyncratic mapping during the tests of left/right responses onto the outcomes for that cue. For example, across test trials with cue D, if participants were biased to choose outcome E as the most likely outcome after a left response and outcome F after a right response, then D-left-E and D-right-F were defined as the full sequences for that cue. Correct classification of a neural pattern from a cue+action trial with D-left would thus occur if the full-sequence classifier guessed D-left-E and if the outcome-only classifier guessed E. Importantly, most participants (23/24) tended toward a particular mapping of responses to outcomes for non-predictive cues, even though there was no basis for this in the statistics of their experience (for the remaining participant, who did not show a response bias for unpredictable trials, we randomly chose a mapping). This subjective definition allowed us to measure accuracy for sequence decoding (**Supplementary Fig. 3a**) and outcome decoding (**Supplementary Fig. 3b**) for trials in which actions did not provide predictive information.

Action decoding

We interpreted sequence decoding in CA–DG as evidence that the cue+action trials elicited a conjunctive representation of the corresponding full cue+action+outcome sequence. Notably, however, the two sequences for each cue always involved different actions (left for one, right for the other). These differential actions were present both in the full-sequence trials used for training and in the cue+action trials used for testing. This leaves open the possibility that left/right button presses or action directions *per se* could underlie sequence decoding in CA–DG. We investigated this possibility in several ways, using both predictable and unpredictable trials.

We examined both CA–DG and the right putamen. The right putamen was included as a control region, to aid in interpreting the results from CA–DG. In particular, it showed sequence decoding in the voxelwise searchlight analysis, but we reasoned that this was because it represented action information rather than sequence information. This reasoning was based on previous studies showing that the right putamen is involved in motor learning[19,20], as well as findings from the current study that — unlike the hippocampus — sequence decoding in the right putamen was unrelated to outcome decoding in V1–V2, both within participants (t_{23}=1.04, P = .31; cf. **Fig. 3a**) and across participants (t(22)=.28, P = .19; cf. **Fig. 3b**). We had no *a priori* hypotheses about the right putamen and do not intend to draw any conclusions about this region. Rather, we used it strictly as a positive control, given that we expected decoding of actions to fail in the hippocampus. That is, such null results would be more readily interpreted if actions could be decoded from another region, by helping rule out alternative explanations (e.g., a problem with the classifier algorithm, an insufficient amount of training or testing examples, etc.).

We first examined action decoding for predictable trials by attempting to cross-classify sequences from different cues. We trained a classifier to discriminate the two sequences for one predictable cue (e.g., A_1) and then tested it on the two sequences for the other predictable cue (A_2). In other words, after training a classifier to discriminate full-sequence trials A_1-left-B_1 and A_1-right-C_1, we tested whether it could decode full-sequence trials A_2-left-B_2 and A_2-right-C_2 (**Supplementary Fig. 4a**). Because cues and outcomes differed

across training and testing sets, the classifier was forced to rely on action information. Left vs. right actions could not be decoded in this way from either CA–DG (t_{23}=−0.39, P = .70) or the right putamen (t_{23}=1.50, P = .15). This across-cue effect in CA–DG was weaker than within-cue sequence decoding in CA–DG (t_{23}=2.27, P = .03). We also tested whether these full-sequence classifiers for one cue could decode cue+action trials for the other cue (A_2-left and A_2-right), since they also preserved the action mappings (**Supplementary Fig. 4b**). Action decoding was again not reliable in CA–DG (t_{23}=1.66, P = .11), though was no longer reliably weaker than within-cue sequence decoding (t_{23}=0.80, P = .43). The right putamen showed marginal action decoding (t_{23}=2.05, P = .05).

These results show that actions were not sufficient for sequence decoding in CA–DG. We next tested whether they were necessary, by examining classification when actions were equated (Supplementary Fig. 5). We trained classifiers to discriminate full-sequence trials with different cues and outcomes but with the same left or right action (for example, A1-left-B1 vs. A2-left-B2). In CA–DG, these within-action classifiers reliably decoded the corresponding cue + action trials with different cues but the same actions during testing (t_{23} = 2.20, P = 0.04). By comparison, such decoding failed in the right putamen (t_{23} = −0.65, P = 0.52). It is difficult to know what is driving the effect in CA–DG, as both the cues and any retrieved sequence information differ, but these findings establish that CA–DG is sensitive to more than the predictive action.

There is a potential issue with using predictable trials to isolate the role of actions in sequence decoding. Namely, the actions and outcomes are perfectly correlated in the full-sequence trials used for classifier training. Thus, decoding of cue + action trials might fail not because actions per se are not represented in CA–DG but rather because the classifier learned to rely on outcome information alone. The fact that action decoding failed in CA–DG when the classifier was tested on full-sequence trials is inconsistent with this possibility, as the outcomes were present. Nevertheless, another approach is to use the unpredictable trials, because the actions were orthogonal to the outcomes.

We trained classifiers to discriminate unpredictable full-sequence trials with left vs. right actions and tested them on unpredictable cue + action trials. Because each action was equally likely to be followed by either outcome, only action information distinguished between classes (Supplementary Fig. 4c). For example, cue D produced four sequences with roughly equal frequency (D-left-E, D-left-F, D-right-E and D-right-F) and we trained the classifier to distinguish the two sequences with the same action from the two with the other action (D-left-E + D-left-F vs. D-right-E + D-right-F). The cue was always identical, and therefore uninformative, and both outcomes appeared on either side of the classification boundary. Consistent with findings from predictable trials, left vs. right actions could not be decoded in CA–DG (t_{23} = 0.11, P = 0.92). By comparison, action decoding was reliable in the right putamen (t_{23} = 2.85, P = 0.009).

Overall, the results in this section are incompatible with the possibility that the observed sequence decoding in the hippocampus was confounded by actions. That is, although the two sequences for a cue had unique actions, we consistently failed to find evidence that these actions per se could be decoded from CA–DG. This stands in contrast to the right

putamen, which similarly showed sequence decoding but also some evidence of action decoding.

Brain-behavior correlations

To examine the behavioral significance of sequence decoding in the hippocampus and outcome decoding in the visual cortex, we related individual differences in classifier accuracy to behavioral performance on the tests outside of the scanner. We calculated the Pearson correlation across participants between mean test RT (accuracy was at ceiling) and either mean sequence decoding accuracy in CA–DG or mean outcome decoding accuracy in V1–V2. Correlations were calculated separately for predictable and unpredictable conditions (Supplementary Fig. 6).

Cross-classification

The logic of our hypotheses and analyses led us to predict that sequence decoding would be successful in the hippocampus and that outcome decoding would be successful in early visual cortex. However, this logic does not necessarily lead to the negative prediction that sequence decoding would completely fail in early visual cortex and that outcome decoding would likewise fail in the hippocampus. Indeed, neural patterns in visual cortex should contain information about the outcome in full-sequence trials because the outcome was present as a stimulus, and thus a classifier trained on full-sequence trials might be able to decode predicted outcomes in visual cortex on cue + action trials. Moreover, a conjunctive representation of the full sequence may be retrieved in the hippocampus by the outcome in outcome-only trials, and thus a classifier trained on outcome-only trials might be able to decode pattern-completed sequences in the hippocampus on cue + action trials. However, we expected that any such decoding would be weak. A full-sequence classifier trained in early visual cortex would have access to neural representations of the different outcome stimuli across sequences, but the discriminability of these outcome representations would be reduced by the addition of a common neural pattern for the shared cue stimulus in these sequences, which is not true of outcome-only trials. An outcome-only classifier trained in the hippocampus might learn something about the corresponding conjunctive representations, but the outcomes provide a minimal retrieval cue (compared to the cue, action and outcome present on full-sequence trials).

Nevertheless, it remains possible that some limited outcome information is present in visual cortex on full-sequence trials and sequence information in the hippocampus on outcome-only trials, just not enough to enable generalization to cue + action trials during testing. To more fully characterize the information available to each classifier, we performed a cross-classification analysis of all trial types. Consistent with the presence of outcome information on full-sequence trials, the full-sequence classifier reliably decoded outcome-only trials in both CA–DG and V1–V2 (Supplementary Fig. 7a). Similarly, the outcome-only classifier reliably decoded full-sequence trials in V1–V2, though not in CA–DG (Supplementary Fig. 7b). This latter result is inconsistent with the possibility that outcomes induced pattern completion in the hippocampus.

In addition to better characterizing the information available to each classifier, the cross-classification analysis affirmed that both the training and testing sets contributed to successful decoding. For example, as evidence of the robustness of our sequence decoding findings, we obtained the same results—that is, reliable classification in CA–DG but not V1–V2—when we swapped the training and testing sets (training on cue + action trials and testing on full-sequence trials) (Supplementary Fig. 7c). However, when we trained on cue + action trials and tested on outcome-only trials (swapping training and testing sets for outcome decoding), the pattern of results was the same but not as statistically reliable. The more robust classification from outcome-only to cue + action than from cue + action to outcome-only in V1–V2 may be partly explained by the fact that we had fewer cue + action trials than outcome-only trials for classifier training. This was intentional, as the experimental design was optimized to maximize the amount of training data for our full-sequence and outcome-only classifiers. Another more general explanation is that training and testing sets should only be interchangeable if they contain equally robust information. However, the outcome-only and cue + action trials may have evoked neural representations of the outcomes that differed in strength, with the former being driven by an external stimulus and the latter reflecting an internal expectation. Because the classifier weights are learned based on the training data and fixed for the testing data, stronger representations on outcome-only trials may have enabled learning of a better boundary in the classifier.

To summarize the cross-classification analyses: across all training/testing permutations, decoding was reliable in the hippocampus only when full-sequence trials were part of either the training or testing set and in early visual cortex only when outcome-only trials were part of either the training or testing set. These results are consistent with our approach of treating full-sequence trials as the best probe of conjunctive representations in the hippocampus and outcome-only trials as the best probe of sensory representations in early visual cortex.

Timecourse analysis

We used a multivariate connectivity approach to examine the temporal dynamics of sequence decoding and outcome decoding (Supplementary Fig. 9). As with earlier analyses, we first trained a classifier in CA–DG on the pattern of beta parameters from the GLM for each full-sequence trial and another classifier in V1–V2 on the GLM parameters for each outcome-only trial. However, instead of testing on the GLM parameters for each cue + action trial (a gamma-weighted average of activity over time), we used z-scored raw activity patterns at various time points in the trial. This allowed us to examine the relationship between classifier accuracy at different TRs across regions. We focused around the peak of the hemodynamic response, isolating the 3rd (4.5 s) and 4th (6 s) TRs after trial onset. Because there were always at least two TRs between trial onsets, each isolated TR was unique to a particular trial (i.e., the 4th TR of one trial always preceded the 3rd TR of the next trial). Within and across these two time points, we examined the relationship over trials of sequence information in the hippocampus and outcome information in visual cortex. If the former precedes the latter, sequence decoding in CA–DG at TR 3 should predict outcome decoding in V1–V2 at TR 4, but V1–V2 outcome decoding at TR 3 should not predict CA–DG sequence decoding at TR 4. We used multinomial regression on the classifier accuracies at each TR across trials to measure this predictive relationship.

Statistics

All tests were evaluated against a two-tailed $P < 0.05$ level of significance. Data collection and analysis were not performed blind to the conditions of the experiment. Because we had no expectation of hemispheric differences, we averaged classifier accuracies across left and right ROIs for all analyses. Classifier accuracy for sequence decoding and outcome decoding did not reliably differ between hemispheres in either the hippocampus or visual cortex ($P > 0.06$). A repeated-measures ANOVA was used to test the key interaction of classifier (sequence decoding vs. outcome decoding) by region (hippocampus (average of CA2–CA3–DG, CA1 and subiculum) vs. early visual cortex (average of V1 and V2)). One-sample t-tests were used to compare classifier accuracies to chance.

To assess within- and across-participant relationships between hippocampal sequence decoding and visual cortex outcome decoding with greater power and fewer comparisons, we pooled CA2–CA3–DG and CA1 into a single CA–DG ROI, and V1 and V2 into a single V1–V2 ROI. For the within-participant relationship between sequence decoding and outcome decoding, we categorized trials as “correct’ if the classifier output in both left and right hemispheres matched the corresponding full sequence and as “incorrect’ if the classifier in either hemisphere produced mismatching output. The across-participant relationship was assessed using Pearson correlation. These same pooled ROIs were also used for several control analyses.

Because classifier accuracies for sequence decoding and outcome decoding met parametric assumptions, such as normality and independence, we used standard parametric tests for primary and control analyses. However, to verify that our results did not depend on such assumptions, we repeated these statistical analyses using a random-effects form of bootstrap resampling[41] (Supplementary Figs. 2 and 8). For each test, we sampled with replacement from the 24 participants 10,000 times. All effects that were reliable with parametric tests were reliable in these non-parametric tests as well. A Supplementary methods checklist is available.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH grants F32 EY023162, S10 OD016277, and R01 EY021755. The authors thank P. Kok and A. Schapiro for helpful comments on an earlier draft.

References

1. Rao RPN, Ballard DH. Nat. Neurosci. 1999; 2:79–87. [PubMed: 10195184]
2. Friston K. Phil. Trans. R. Soc. Lond. B. 2005; 360:815–836. [PubMed: 15937014]
3. Clark A. Behav. Brain. Sci. 2013; 36:181–204. [PubMed: 23663408]
4. Smith FW, Muckli L. Proc. Natl. Acad. Sci. USA. 2010; 107:20099–20103. [PubMed: 21041652]
5. Alink A, Schwiedrzik CM, Kohler A, Singer W, Muckli L. J. Neurosci. 2010; 30:2960–2966. [PubMed: 20181593]
6. Schapiro AC, Kustner LV, Turk-Browne NB. Curr. Biol. 2012; 22:1622–1627. [PubMed: 22885059]

7. Hawkins, J.; Blakeslee, S. Times Books: 2004
8. Kok P, Jehee JF, de Lange FP. Neuron. 2012; 75:265–270. [PubMed: 22841311]
9. Marr D. Phil. Trans. R. Soc. B. 1971:23–81. [PubMed: 4399412]
10. Cohen, NJ.; Eichenbaum, H. MIT Press: 1993
11. Leutgeb S, Leutgeb JK. Learn. Mem. 2007; 14:745–757. [PubMed: 18007018]
12. Ji D, Wilson MA. Nat. Neurosci. 2007; 10:100–107. [PubMed: 17173043]
13. Bosch SE, Jehee JF, Fernández G, Doeller CF. J. Neurosci. 2014; 34:7493–7500. [PubMed: 24872554]
14. Hindy NC, Turk-Browne NB. Cereb. Cortex. 2015
15. O'Doherty J, et al. Science. 2004; 304:452–454. [PubMed: 15087550]
16. Bar M, et al. Proc. Natl. Acad. Sci. USA. 2006; 103:449–454. [PubMed: 16407167]
17. Dale AM. Hum. Brain Mapp. 1999; 8:109–114. [PubMed: 10524601]
18. Yushkevich PA, et al. Hum. Brain Mapp. 2015; 36:258–287. [PubMed: 25181316]
19. Aly M, Turk-Browne NB. Cereb. Cortex. 2015
20. Duvernoy HM. Springer: 2005.
21. Carr VA, Rissman J, Wagner AD. Neuron. 2010; 65:298–308. [PubMed: 20159444]
22. Treves A, Tashiro A, Witter MP, Moser EI. Neuroscience. 2008; 154:1155–1172. [PubMed: 18554812]
23. Ketz N, Morkonda SG, O'Reilly RC. PLoS Comput. Biol. 2013; 9:1–9.
24. Nakashiba T, et al. Cell. 2012; 149:188–201. [PubMed: 22365813]
25. Duncan K, Tompary A, Davachi L. J. Neurosci. 2014; 34:11188–11198. [PubMed: 25143600]
26. Bakker A, Kirwan CB, Miller M, Stark CE. Science. 2008; 319:1640–1642. [PubMed: 18356518]
27. Fischl B, et al. Cereb. Cortex. 2008; 18:1973–1980. [PubMed: 18079129]
28. Hinds OP, et al. NeuroImage. 2008; 39:1585–1599. [PubMed: 18055222]
29. Dale AM, Fischl B, Sereno MI. NeuroImage. 1999; 9:179–194. [PubMed: 9931268]
30. Wang L, Mruczek RE, Arcaro MJ, Kastner S. Cereb. Cortex. 2014
31. Kriegeskorte N, Goebel R, Bandettini P. Proc. Natl. Acad. Sci. USA. 2006; 103:3863–3868. [PubMed: 16537458]
32. Debas K, et al. NeuroImage. 2014; 99:50–58. [PubMed: 24844748]
33. Gabitov E, Manor D, Karni A. J. Cogn. Neurosci. 2015; 27:736–751. [PubMed: 25390206]
34. Smith SM, et al. NeuroImage. 2004; 23(Suppl 1):S208–S219. [PubMed: 15501092]
35. Greve DN, Fischl B. NeuroImage. 2009; 48:63–72. [PubMed: 19573611]
36. Jenkinson M, Bannister P, Brady M, Smith S. NeuroImage. 2002; 17:825–841. [PubMed: 12377157]
37. Andersson JL, Jenkinson M, Smith S. FMRIB Tech. Rep. 2007; TR07JA2
38. O'Reilly RC, Rudy JW. Psychol. Rev. 2001; 108:311. [PubMed: 11381832]
39. Komorowski RW, Manns JR, Eichenbaum H. J. Neurosci. 2009; 29:9918–9929. [PubMed: 19657042]
40. Hindy NC, Solomon SH, Altmann GTM, Thompson-Schill SL. Cereb. Cortex. 2015; 25:884–894. [PubMed: 24127425]
41. Efron B, Tibshirani R. Stat. Sci. 1986:54–75.

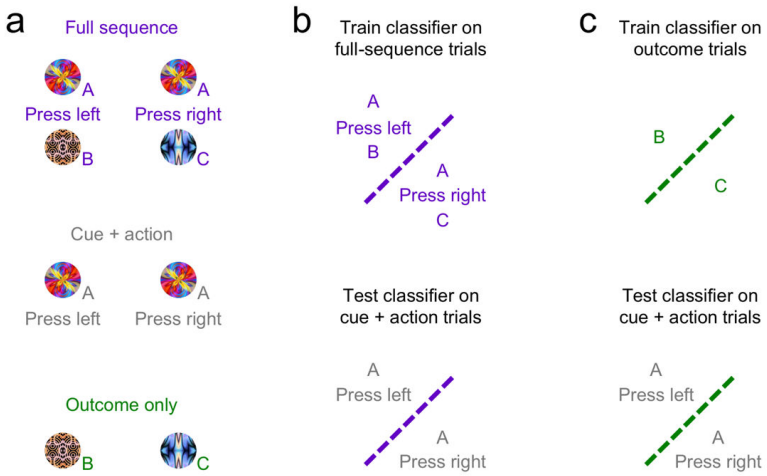

Figure 1.

Analysis approach. **(a)** There were three types of trials during fMRI: full-sequence trials (purple lettering), in which cue A was replaced by outcome B if a button was pressed with the left hand and by outcome C if a button was pressed with the right hand; cue+action trials (gray), in which A was replaced by a blank screen upon either button press; and outcome-only trials (green), in which B or C appeared in isolation without a button press. **(b)** Pattern completion was operationalized as the amount of neural evidence elicited by a cue and action about the corresponding full sequence. This evidence was measured with a multivariate classifier trained on full-sequence trials to distinguish the two sequences for each cue and tested on cue+action trials (sequence decoding). **(c)** Predictive coding was operationalized as the amount of neural evidence elicited by a cue and action about the expected outcome. This evidence was measured with a multivariate classifier trained on outcome-only trials to distinguish the two outcomes for each cue and tested on cue+action trials (outcome decoding). For both sequence decoding and outcome decoding, separate classifiers were trained and tested on patterns of BOLD activity in ROIs from the hippocampus and visual cortex.

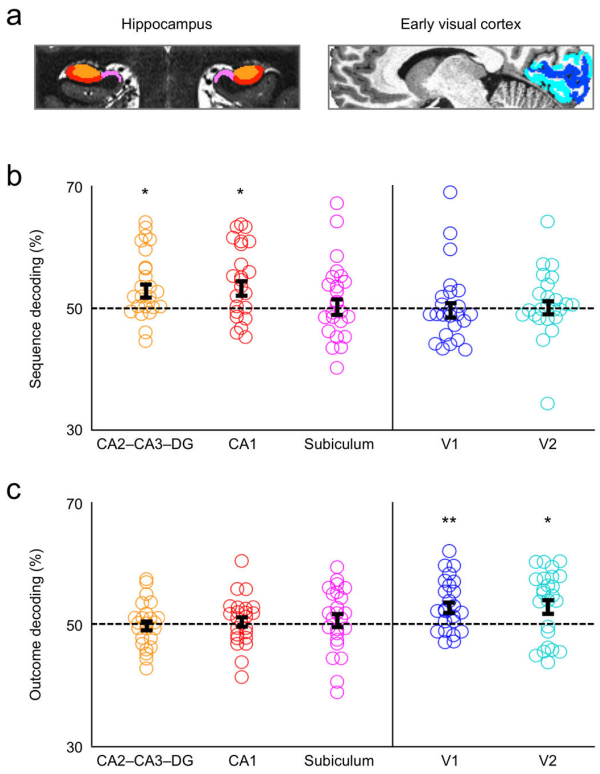

Figure 2.

Decoding performance. **(a)** *A priori* ROIs included CA2–CA3–DG, CA1, and subiculum in the hippocampus, and V1 and V2 in early visual cortex. **(b)** Sequence decoding was reliable in CA2–CA3–DG (t_{23}=2.53, P = .02) and CA1 (t_{23}=2.72, P = .01), but not in subiculum, V1, or V2 (Ps>.81). **(c)** Outcome decoding was reliable in V1 (t_{23}=3.17, P = .004) and V2 (t_{23}=2.51, P = .02), but not in CA2–CA3–DG, CA1, or subiculum (Ps>.57). Error bars depict ±1 s.e.m. *P<.05, **P<.01

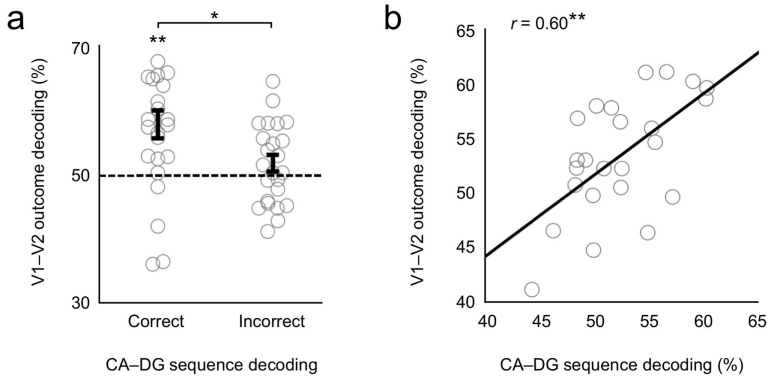

Figure 3.

Hippocampal-visual relationship. **(a)** Outcome decoding in V1–V2 was more reliable (t_{23}=2.32, P = .03) for trials on which sequence decoding in CA–DG was correct (vs. 50% chance: t_{23}=3.45, P = .002) vs. incorrect (t_{23}=1.05, P = .30). Error bars depict ±1 s.e.m. **(b)** Individual differences in V1–V2 outcome decoding could be predicted from CA–DG sequence decoding (robust r_{22}=.60, P = .002). *P<.05, **P<.01