Cerebral Cortex, May 2016;26: 1853–1865

doi: 10.1093/cercor/bhv030
Advance Access Publication Date: 9 March 2015
Feature Article

FEATURE ARTICLE

Action-Based Learning of Multistate Objects in the Medial Temporal Lobe

Nicholas C. Hindy[1] and Nicholas B. Turk-Browne[1,2]

[1]Princeton Neuroscience Institute, and [2]Department of Psychology, Princeton University, Princeton, NJ 08544, USA

Address correspondence to Nicholas Hindy, Peretsman-Scully Hall 324, Princeton, NJ 08544, USA. Email: nhindy@princeton.edu

Abstract

Actions constrain perception by changing the appearance of objects in the environment. As such, they provide an interactive basis for learning the structure of visual input. If an action systematically transforms one stimulus into another, then these stimuli are more likely to reflect different states of the same persisting object over time. Here we show that such multistate objects are represented in the human medial temporal lobe—the result of a mechanism in which actions influence associative learning of how objects transition between states. We further demonstrate that greater recruitment of these action-based representations during object perception is accompanied by attenuated activity in stimulus-selective visual cortex. In this way, our interactions with the environment help build visual knowledge that predictively facilitates perceptual processing.

Key words: action, medial temporal lobe, object perception, predictive coding, representational similarity

Introduction

Different visual stimuli can signify the same object at different moments in time, such as an open versus closed umbrella, a book on the table versus shelf, or the front versus side view of a face. To recognize and keep track of objects, the identity of each object must be preserved across such differences (Biederman and Gerhardstein 1993; DiCarlo et al. 2012; Hindy et al. 2012, 2015). By way of their invariant tuning properties, high-level areas of the visual system, including inferior temporal cortex and lateral occipital (LO) cortex, may be sufficient for recognizing objects across simple transformations, such as rotated viewpoints or translated locations (Grill-Spector et al. 2001; DiCarlo et al. 2012).

However, generalization of identity across multiple states of an object that have little or no overlap in features (e.g., a fresh laid egg, a painted egg, or an omelet) may require learning the dynamic structure of the object (e.g., that a fresh laid egg can become either a painted egg or an omelet). Accordingly, such multistate object recognition may depend on associative learning mechanisms in the medial temporal lobe (MTL) (Cohen and Eichenbaum 1993; Miyashita 1993). In particular, different states of the same object are often observed contiguously in time, and such temporal regularities induce changes in how stimuli are represented in perirhinal cortex (PRC) and entorhinal cortex (ERC) (Miyashita 1988; Wirth et al. 2003; Schapiro et al. 2012).

In the current study, we investigate how actions influence the formation of multistate object representations in the MTL, beyond what can be gleaned from passively perceiving state transitions. Specifically, actions provide a rich source of information about the relational structure of stimuli that constitute the same object (e.g., that a fresh laid egg can only become a painted egg when painted and an omelet when scrambled). In a more formal example, if stimulus B appears whenever action X is performed on stimulus A, and stimulus C appears whenever a different action Y is performed on A, then A, B, and C may all be states of the same object, linked by action in a tree-like structure. Although the stimulus transitions in this example are weakly probabilistic (if actions are equally frequent: $P[B|A] = P[C|A] = 0.5$), the actions add predictive power, allowing the outcome stimulus to be anticipated deterministically ($P[B|AX] = P[C|AY] = 1.0$). If particular regions of the MTL are sensitive to action information, actions may facilitate generalization of object identity across unique but predictable stimuli by binding object states into structured representations.

Note that this process involves building stimulus–response associations, a function typically ascribed to the striatum rather than the MTL (Poldrack et al. 2001; Yin and Knowlton 2006). However, in the current study, we examine response learning in a different sense—the role that responses play in mediating stimulus–stimulus learning in the MTL. Such learning could occur in an error-driven manner (e.g., predicting B after AY and receiving C instead), but this error signal is an unexpected stimulus rather than a reward or punishment. The role of the striatum in learning from such *stimulus* prediction errors (in the absence of motivational significance) is unclear (Niv and Schoenbaum 2008). Moreover, the MTL is more involved in response learning (Sadeh et al. 2010; Wimmer and Shohamy 2012; Shohamy and Turk-Browne 2013) and prediction error (Henson and Gagnepain 2010; Chen et al. 2011) than previously thought. We test the possibility that actions are encoded by the MTL, leading to stimulus–response–stimulus associations that serve as the foundation for dynamic object representations.

The current study was conducted over 2 days (Fig. 1*A*). On the first day, participants underwent a training regimen in which they learned associations between stimuli that were or were not linked by predictive actions. On each trial, participants were presented with a cue stimulus, pressed a button with either their left or right hand as an action (their choice), and were then presented with an outcome stimulus (Fig. 1*B*). The stimuli were drawn from 4 triads (Fig. 2*A*). Two of the triads belonged to the *predictable* condition: given cue A, outcome B appeared with high probability when the left button was pressed and outcome C appeared with high probability when the right button was pressed (Fig. 2*B*). The other 2 triads belonged to the *unpredictable* condition: given cue D, outcomes E and F were equally likely after either a left or right button press (Fig. 2*C*). Actions were meaningless for unpredictable triads, as they did not provide any information about which outcome would appear. Since unpredictable triads were otherwise identical to predictable triads, they served as a baseline control for the learning of stimulus–stimulus transitions in the absence of predictive action. Participants repeated this exploratory training until they achieved criterion performance in a behavioral test (Fig. 1*D*).

On the second day, participants completed additional training with directed actions, both to refresh the associations and to balance the frequencies of all stimuli and transitions (Fig. 1*C*). For example, if they responded left more than right during the exploratory training, they were more likely to be instructed to respond right in the directed training. Thus, across all training sessions, the stimulus and transition probabilities were identical for both outcome stimuli and for both the predictable and unpredictable conditions. The one thing that differed

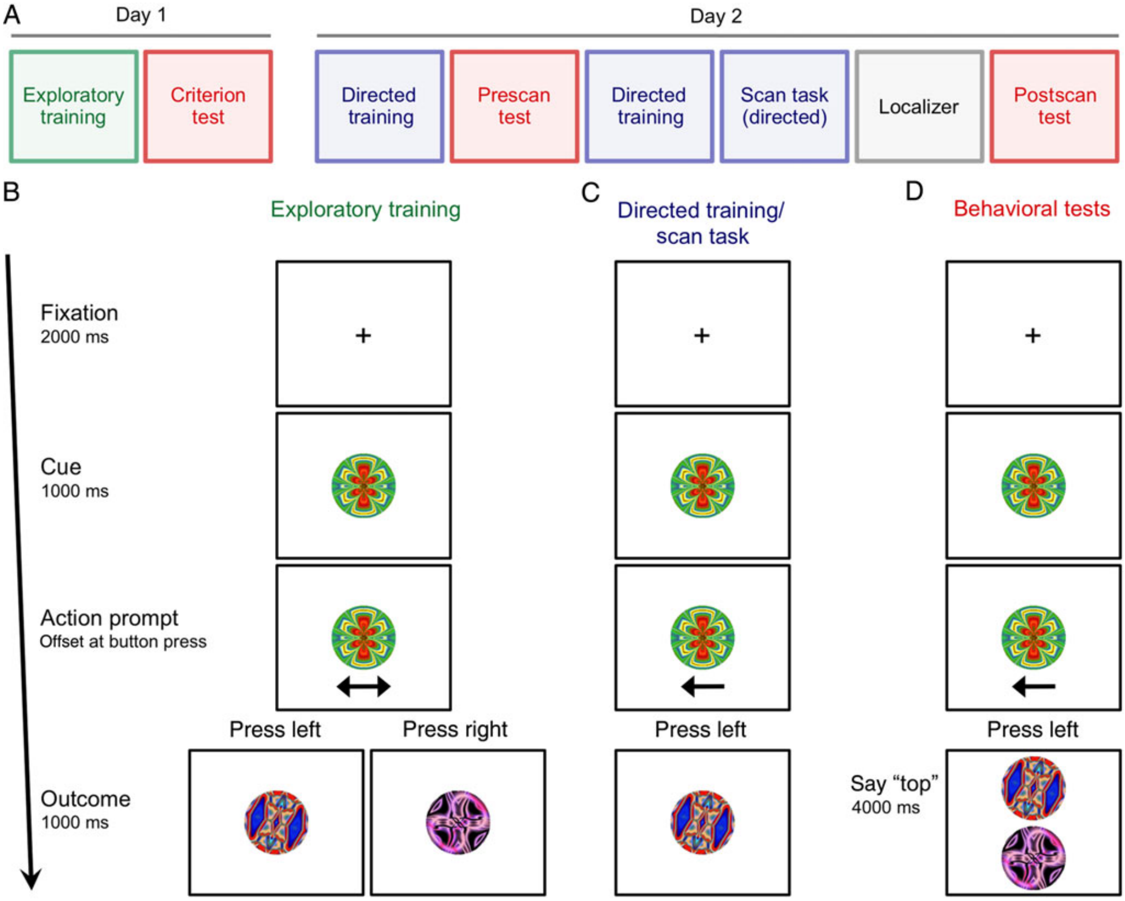

Figure 1. Procedure. (*A*) Study timeline, including exploratory training in which participants chose their own actions (green), tests for whether they had learned cue–action–outcome associations (red), and directed training in which their actions were instructed (blue). (*B*) Trial sequence for the exploratory training. Each trial began with a cue at fixation. A double-headed arrow prompted participants to press a button with their choice of left or right hand, at which point an outcome appeared immediately. (*C*) In the directed training and scan task, a single-headed arrow prompted the desired action, for counterbalancing reasons. (*D*) Behavioral tests were conducted to assess participants' learning at different stages. Each test trial involved a verbal "top" or "bottom" response to select which of 2 outcomes seemed most probable.

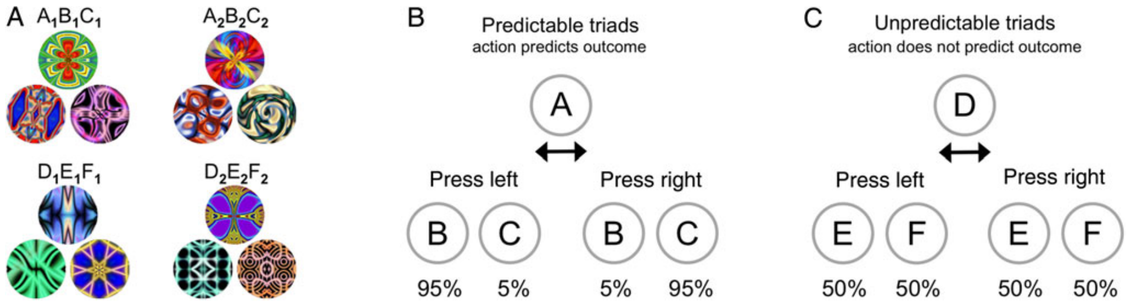

Figure 2. Stimulus triads. (*A*) Two predictable triads and 2 unpredictable triads were each composed of one cue stimulus and 2 outcome stimuli. (*B*) For predictable triads ($A_1B_1C_1$, $A_2B_2C_2$), given the cue (A_1, A_2), one specific outcome (B_1, B_2) appeared with 95% probability when the left button was pressed, and a different specific outcome (C_1, C_2) appeared with 95% probability when the right button was pressed. (*C*) For unpredictable triads ($D_1E_1F_1$, $D_2E_2F_2$), given the cue (D_1, D_2), either of 2 outcomes (E_1 or F_1, E_2 or F_2) appeared with 50% probability when either the left or right button was pressed.

between conditions was whether the action was informative about which outcome would appear. After another behavioral test, participants again performed the directed version of the task in the scanner, while we measured blood oxygen level-dependent (BOLD) activity patterns in the MTL. A localizer scan was included at the end of the scan session to identify stimulus-selective areas of visual cortex, and a final behavioral test was administered to verify that participants had retained the action-based associations.

If predictive actions allow the MTL to learn the multistate structure of an object, then the MTL should treat predictable triads differently from unpredictable triads. In particular, the multistate objects learned from predictable triads should be more recognizable, such that the underlying neural representations of these triads are more distinguishable from one another (Naya and Suzuki 2011). To test this hypothesis, we examined multivariate patterns of BOLD activity in 4 MTL regions-of-interest (ROIs): the hippocampus, PRC, ERC, and parahippocampal cortex (PHC). We used a hierarchical analysis approach, in which we first tested within each ROI for an overall effect of predictability on the representation of the stimulus triads. We found that activity patterns in PRC and ERC for predictable triads were more distinguishable from one another than patterns for unpredictable triads. In these ROIs that showed a difference in triad similarity, we then tested alternative explanations of the underlying representational change: internal merging between states of the same object, or external differentiation between states of different objects. Follow-up analyses examined whether a subjective sense of predictability was sufficient to induce representational changes, and how action-based learning in MTL affected stimulus processing in visual cortex.

Materials and Methods

Participants

Twenty-four individuals (11 female, aged 18–31 years) from the Princeton University community participated in the study. Each participant was right-handed and had normal or corrected-to-normal vision. One additional participant was excluded from data analysis and replaced due to excessive motion in the scanner (moved ~14.5 mm during the scan and partially out of the field of view) and another due to unusually poor performance on the scan task (failed to press the indicated button on >27% of trials, which was 4 SDs worse than the mean). Participants were paid $20 per hour and provided informed consent to a protocol approved by the Princeton University Institutional Review Board.

Stimuli

The primary stimulus set of 12 fractal-like images is displayed in Figure 2A. An additional 48 unique fractal images were used as novel outcomes during the scan task, and 72 additional unique fractal images were used for the localizer. All fractal images were created using ArtMatic Pro (www.artmatic.com), with a subset of the images used in a previous study (Schapiro et al. 2012). Fractal images subtended ~4° of visual angle in diameter on the training/testing laptop computer, and 4.5° in the scanner. We counterbalanced the assignment of images to predictable and unpredictable triads, and randomly assigned them to be cues or outcomes.

Training

Training consisted of two 30-min sessions prior to the scanning session. Within each training session, participants responded to 80 repetitions of each cue stimulus. For each trial, an arrow prompt appeared below the cue after 1000 ms, and the participant pressed the left or right button using the corresponding hand. Immediately upon button press, the cue stimulus was replaced by an outcome stimulus. For predictable triads, one specific outcome (outcome_1) appeared with 95% probability when the left button was pressed and the other outcome (outcome_2) appeared the remaining 5% of the time; when the right button was pressed, outcome_2 appeared with 95% probability and outcome_1 with a 5% probability. For unpredictable triads, outcome_1 or outcome_2 appeared with 50% probability when either the left or right button was pressed. There were 2 motivations for making the predictable triads only 95% (instead of 100%) consistent: (1) to make the task of learning probabilistic relationships credible to participants, and (2) to protect learning against extinction when predictions were occasionally violated during the scan task.

Exploratory (Day 1)

The first training session was an exploratory phase in which the arrow prompt was always a double-headed arrow, and participants decided for each cue stimulus whether to press the left or right button. This session included 320 trials and was conducted on a laptop computer ~24 h prior to scanning, which allowed time for additional training in case a participant did not immediately learn to criterion all of left/right response mappings for the predictable triads. Throughout exploratory training, a response meter at the bottom of the screen tracked the proportion of left and right button presses, and participants were instructed to keep the meter pointer within a specified central zone (as a visual aid to encourage self-counterbalancing).

Directed (Day 2)

After tracking responses during exploratory training, any frequency differences were balanced through directed training. On each trial, a single-headed arrow prompt was shown after the cue to specify the desired left or right response. The directed training session occurred immediately prior to the scan. The first 240 trials of directed training were performed on a laptop computer and the final 80 trials were performed in the scanner during acquisition of structural images (to familiarize participants with their appearance in this new environment). Across the exploratory and directed training sessions, the frequency of stimulus transition was equated for all triads.

Behavioral Tests

To verify learning of the predictable action-based associations, each participant performed 2 prescan tests and one postscan test. On each test trial, a cue stimulus appeared at fixation. Below the cue, a single-headed arrow pointed left or right, and participants were instructed to press the corresponding button. The cue and arrow disappeared, replaced by the 2 possible outcomes for that cue, presented above and below where the cue had been. One outcome correctly completed the cue–action–outcome sequence, while the other outcome completed the cue–action–outcome sequence for the other action. Each test included 16 predictable and unpredictable trials in random order, with 2 trials of each cue–action–outcome sequence (4 trials for each cue). Participants spoke aloud either “top” or “bottom” to indicate which outcome was expected given the cue and action. Verbal responses were used to avoid the button presses used for our trained actions. The only feedback that participants received regarding their test performance was their accuracy for predictable triads on the first prescan test. Participants were required to be 100% accurate on the first prescan test in order to ensure action-based learning of predictable triads during the exploratory training phase. For analyses based on test consistency, consistent and inconsistent triads were coded post hoc for each participant based on performance on all 3 tests.

Scan Task

The task in the scanner resembled the training tasks, and included 336 trials equally distributed across 6 runs, with each run ~6 min in duration. Upon beginning the scan, participants knew to expect a final behavioral test after the scan in order to measure how their knowledge about the action-based relationships had developed over the course of the scan task. Specifically, participants were instructed: “This is a test of how learning affects perception. For each trial, a fractal will appear at the center of the screen. When a single-headed arrow appears, press left or right as directed to see the next fractal. Continue to keep track of probabilistic relationships between button presses and fractal pairs.” As during training, each trial in the scanner included 3 parts: a cue stimulus for 1000 ms, an action prompt, and an outcome stimulus for 1000 ms. Identical to the directed training phase, the action prompt was a single-headed arrow pointing left or right below the cue, which remained on the screen until a button press or until a 1500 ms response window elapsed. Using a separate response box for each hand, participants pressed the left or right button that corresponded to the direction that the arrow was pointing. If participants did not press a button within the response window, the cue stimulus and action prompt were replaced with a fixation cross that remained on the screen until the subsequent trial. The average length of fixation between trials was 3612 ms, which included an interval of 1500, 3000, or 4500 ms of fixation, along with additional fixation determined by the response time for the previous trial. The order of trial types within each run was optimized for statistical power using optseq2 (Dale 1999; http://surfer.nmr.mgh.harvard.edu/optseq/).

During the scan task, 50% of trials included a predictable cue and 50% included an unpredictable cue. Outcome stimuli were likewise presented with equal frequency across predictable and unpredictable conditions. In total, predictable triads were 28.6% of all trials and unpredictable triads were 35.7% of all trials. There were fewer trials for predictable triads because an additional 7.1% of trials with predictable cues involved presenting the incorrect, counter-predicted outcome (i.e., as if the other action had been performed). Other trial types were also interspersed throughout the scan task in order to probe repetition-related univariate activity in LO and MTL (Supplementary Fig. S1). Specifically, 28.6% of trials included a novel-outcome stimulus that was not seen at all during training, and 28.6% of trials included an unassociated outcome that had been paired with a different cue during training (counterbalanced across all cues).

Localizer

The final scanning run was a functional localizer to identify stimulus-selective LO (Kourtzi and Kanwisher 2000). During this run, participants viewed 16 s blocks of fractal images, alternating with fixation and with blocks of phase-scrambled versions of the images. Each stimulus was presented for 1000 ms followed by 500 ms fixation between trials, and participants performed a one-back repetition detection task. The localizer run was ~7 min in duration, and included 72 unique fractal images, along with 72 phase-scrambled versions of those images. The images were divided into groups of 9, spread across 8 blocks of fractal images and 8 blocks of scrambled images, along with 1 repeated image within each block. Supplementary Figure S2 displays example stimuli and an overlay map of individual LO ROIs in each hemisphere.

Data Acquisition

Structural and functional MRI data were collected on a 3 T Siemens Skyra system with a 16-channel head coil. Structural data were acquired with a high-resolution T_1-weighted MPRAGE sequence, while functional data were acquired with a T_2*-weighted echo-planar imaging sequence with a voxel size of 3 × 3 × 4 mm (TR = 1500 ms, TE = 32 ms). A B0 fieldmap was collected at the end of the experiment. Twelve seconds preceded data acquisition in each functional run to approach steady-state magnetization.

MTL Segmentation

MTL regions, including hippocampus, PRC, ERC, and PHC, were defined probabilistically in MNI space based on a database of manual MTL segmentations from a separate set of 24 participants. Manual segmentations were created on T_2-weighted TSE images using anatomical landmarks (Duvernoy 2005; Carr et al. 2010; Schapiro et al. 2012), and then registered to an MNI template. Voxels in the MNI template were assigned to ROIs based on a probability threshold of $P > 0.5$. Nonlinear registration (FNIRT; Andersson et al. 2007) was used to register each participant's high-resolution MPRAGE to the probabilistic atlas of MTL in MNI space. Supplementary Figure S3 displays the MTL ROIs for PRC, ERC, and hippocampus in the native high-resolution space of each participant.

Data Analysis

Software

Preprocessing and GLM analyses were performed in FSL5 (Smith et al. 2004), and the resulting parameter estimates were processed further in MATLAB for subsequent analyses.

Preprocessing

Data were corrected for slice-acquisition time and head motion, high-pass temporally filtered (using a 50 s period cutoff for event-related runs, and a 128 s period cutoff for the blocked localizer run), and spatially smoothed using a 5 mm FWHM Gaussian kernel. Such spatial smoothing has been shown to benefit multivariate analysis (Op de Beeck 2010). Regardless, the same pattern of results was found for unsmoothed data (Supplementary Fig. S4). Functional runs were registered to each participant's high-resolution MPRAGE image using FLIRT boundary-based registration with B0-fieldmap correction, and then normalized using FNIRT to a standard MNI template with 2 mm isotropic voxels (Andersson et al. 2007; Greve and Fischl 2009).

General Linear Modeling

Parameter estimates of BOLD response amplitude were computed using FILM, with a modified GLM that included temporal autocorrelation correction and 6 motion parameters as nuisance covariates. Each trial was modeled with a boxcar that matched the average trial duration for the participant (between 2500 and 2600 ms, depending on the participant's mean response time) and then convolved with a double-gamma hemodynamic response function. Three separate GLMs were used to examine: (1) triad similarity, (2) within-triad and across-triad similarity, and (3) trial-by-trial MTL–cortical interaction.

In GLM_1, which was used to examine triad similarity, each of the 4 triads was modeled with a single regressor and its temporal derivative. In GLM_2, which was used to examine within-triad and across-triad similarity (including the subjective predictability analyses), each cue–action–outcome sequence was modeled with its own regressor and temporal derivative. This resulted in 12 regressors of interest: 4 regressors for the cue–action–outcome sequences of the 2 predictable triads (e.g., A_1-left-B_1 and A_1-right-C_1), and 8 regressors for the cue–action–outcome sequences of the 2 unpredictable triads (e.g., D_1-left-E_1, D_1-left-F_1, D_1-right-E_1, and D_1-right-F_1). In both GLM_1 and GLM_2, parameter estimates were calculated within run and then averaged across runs for each participant and condition. Also, both GLMs included additional regressors and their temporal derivatives for: counter-predicted, novel-outcome, and unassociated-outcome trials, as well as trials for which the participant failed to press the indicated button. In GLM_3, which was used to examine trial-by-trial MTL–LO interactions, each of the 56 trials in every run was modeled with its own regressor and temporal derivative.

Similarity-Based Analyses

Pattern similarity was measured as the Pearson correlation across voxels between parameter estimates for triads from GLM_1, cue–action–outcome sequences from GLM_2, or trials from GLM_3. For each similarity analysis, vectors of parameter estimates were z-scored within voxel across triads, sequences, or trials, respectively. Within-voxel normalization was performed in order to remove common activation in response to the highly similar fractal stimuli, as well as control for potential differences in the univariate signal magnitude between conditions (Sayres and Grill-Spector 2008). The same basic pattern of results was found before parameter estimates were normalized (Supplementary Fig. S4).

Triad similarity for predictable triads was measured as the correlation between the average vectors for $A_1B_1C_1$ and $A_2B_2C_2$, while triad similarity for unpredictable triads was measured as the correlation between the average vectors for $D_1E_1F_1$ and $D_2E_2F_2$. In each case, this overall similarity between pairs of triads was a measure of the neural separation of the triads with respect to one another, with lower similarity interpreted as greater separation.

Sequence similarity for every triad type was measured as the correlation between cue–action–outcome sequences from within the same triad (e.g., A_1-left-B_1 and A_1-right-C_1) or across different triads of the same condition (e.g., A_1-left-B_1 and A_2-right-C_2). To control for multiple comparisons, primary analyses of within-triad and across-triad similarity focused exclusively on MTL subfields that showed a triad-similarity difference between conditions. We reasoned that it only makes sense to ask follow-up questions about this difference (e.g., how it manifests in representational space) if the difference exists in the first place.

To examine the effect of subjective predictability on sequence similarity, we used performance on behavioral tests to label each triad as predictable-consistent, unpredictable-consistent, or unpredictable-inconsistent. For unpredictable-consistent and unpredictable-inconsistent triads, we focused exclusively on trials in which participants subjectively expected the outcome that appeared. That is, we considered only those trials in which the cue–action–outcome sequence matched the participant's preferred left/right response mapping from the behavioral tests. This is most obvious for the unpredictable-consistent triads, for which one of the outcomes was chosen with 100% consistency across all the tests for a particular cue and action: we restricted analysis to the half of scanned trials for this cue and action that contained this outcome. However, this also applies to unpredictable-inconsistent triads for which one of the left/right response mappings was indicated more often than the other left/right response mapping. For instance, a participant may have associated a left button press with outcome E on 83% of test trials, and with outcome F on 17% of test trials. In this case, we again restricted analysis to the half of scanned trials for this cue and action that contained the majority outcome. When an unpredictable-inconsistent triad did not have a preferred mapping (i.e., there was exactly 50% consistency), both cue–action–outcome sequences were included in the analysis.

Triad reactivation for every triad type was measured as the Pearson correlation between the pattern of activity evoked by the cue–action–outcome sequence on each trial and the average activity pattern across all cue–action-outcome sequences from the same triad. For example, triad reactivation (Y) for trial *i* of predictable triad *j* can be formalized as:

$$Y_i = r\left(Pattern_i, \left(\sum_{1}^{n} A_jB_j + \sum_{1}^{m} A_jC_j\right)/(n+m)\right),$$

where *n* is the total number of A_jB_j trials and *m* is the total number of A_jC_j trials. To examine possible interactions between the MTL and LO, we related this measure of triad reactivation to the level of activity in LO on the same trial using linear regression. The same procedure was also used over the whole brain, relating trial-by-trial MTL triad reactivation to the activity of every voxel. Activity maps for these analyses were corrected for multiple comparisons across the whole brain at $P < 0.05$, based on a voxelwise α of $P < 0.001$ and a cluster-forming threshold of 25 voxels determined using 3dClustSim (Cox 1996; http://afni.nimh.nih.gov/pub/dist/doc/program_help/3dClustSim.html).

All correlation coefficients were Fisher transformed prior to statistical analysis. For all analyses, pattern similarity was calculated separately for each hemisphere and then averaged across hemispheres to reduce multiple comparisons. We used paired-sample *t*-tests to compare similarity for objectively predictable and unpredictable triads. For analyses that depended on subjective predictability, we used subject-level bootstrap resampling (Efron and Tibshirani 1986) to assess random-effects reliability. This approach can be useful when the number of items varies across participants as a result of post hoc coding from behavioral responses (Kim et al. 2014). For each bootstrap, we standardized similarity and activity measures within participant and then sampled with replacement from the 24 participants 10 000 times. If a sampled participant did not have any trials of a particular trial type (either Up or Ui), this participant did not contribute any trials of that trial type to the bootstrap iteration while nonetheless contributing to other trial types. (However, each Up and Ui bootstrap statistic was equally reliable when resampling was constrained to participants with trials of that particular type.) The intuition behind the subject-level bootstrap approach used here is that insofar as one or a small number of participants is driving the effect (hence, it is unreliable), a large proportion of these samples will miss these participant(s) and show little evidence of the effect. If the participants are substitutable in terms of their influence when sampled, then the effect will be highly stable and reliable. This technique provides significance values (and confidence intervals) in terms of proportion of resamples for each parameter estimate, or difference between parameter estimates, below some threshold. By *z*-scoring values within each participant prior to pooling, we eliminated between-subject variance, thus allowing us to test for within-subject effects.

Results

Behavior

Participants were required to be 100% accurate for predictable triads on the criterion test immediately following exploratory training on Day 1. Only 2 participants did not achieve this performance level after one round of training; they completed one half of the training again and achieved 100% test accuracy. On Day 2, accuracy for predictable triads was 98.4% on average (SD = 4.2%) in the prescan test and 99.5% on average (SD = 2.6%) in the postscan test (both means above chance of 50%, $Ps < 0.001$).

Although both outcomes were equally likely for unpredictable triads, participants varied in terms of how consistently they chose a particular outcome given a cue and action at test. For example, after cue D and a left action, some participants reliably chose outcome E, others reliably chose outcome F, and others chose both E and F on different test trials or sessions. In later analyses, we exploit these subjective (and spurious) associations.

During the scan task, participants pressed the indicated left- or right-hand button on 98.2% of trials (SD = 2.4%), with a mean response time of 589 ms (SD = 91 ms). Neither the accuracy ($P = 0.27$) nor the response time ($P = 0.53$) of button presses reliably differed between predictable and unpredictable triads.

Object Learning

Prior studies of associative learning suggest that the formation of object representations in the MTL is reflected as changes in the similarity among stimulus-evoked patterns of BOLD activity (Schapiro et al. 2012). Does learning the tree-like structure of predictable triads, in which A transitions to B after a left button press and to C after a right button press, lead these stimuli to be bound together as a single object? To test for such action-based learning in the MTL, we calculated the Pearson correlation of BOLD activity patterns over voxels for different cue–action–outcome events.

Our first and broadest prediction concerned object learning evidenced by neural separation between triad representations. To determine the similarity of the predictable triads and unpredictable triads with respect to each other, we measured the correlation of the average patterns for the 2 triads within each condition (Fig. 3*A*). We reasoned that if predictable triads come to be represented as objects, then MTL patterns of activity for predictable triads (i.e., $A_1B_1C_1$ vs. $A_2B_2C_2$) should be less correlated than patterns for unpredictable triads (i.e., $D_1E_1F_1$ vs. $D_2E_2F_2$). To measure how each triad was represented in an MTL ROI (Fig. 3*B*), we calculated the average voxelwise pattern of activity across all cue–action–outcome sequences of the same triad. For example, the pattern for $A_1B_1C_1$ was calculated as the average of the patterns for A_1B_1 and A_1C_1. This produced 4 averaged triad patterns per participant and ROI: $A_1B_1C_1$, $A_2B_2C_2$, $D_1E_1F_1$, and $D_2E_2F_2$.

Comparing the similarity of these patterns within condition (Fig. 3*C*), we found significantly lower correlations for predictable compared with unpredictable triads in both PRC ($t_{(23)} = -4.87$, $P < 0.001$) and ERC ($t_{(23)} = -3.75$, $P = 0.001$; no differences were

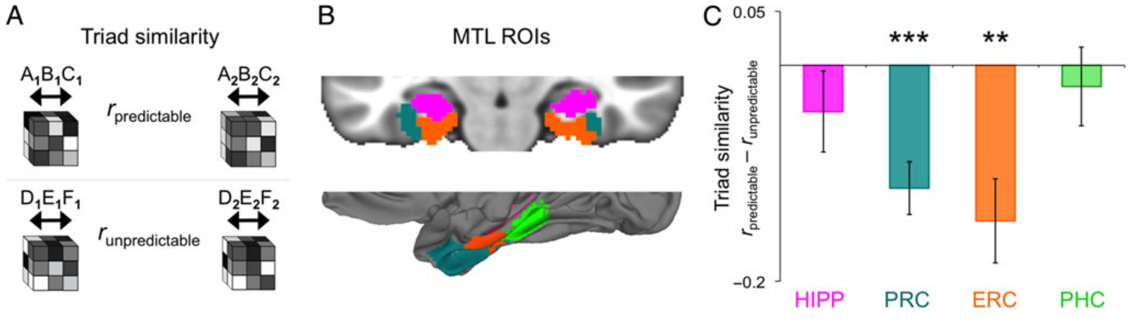

Figure 3. Triad similarity. (*A*) After averaging across all trials for each triad, we calculated the pattern similarity over voxels within each MTL ROI between the 2 predictable triads and between the 2 unpredictable triads. (*B*) MTL ROIs, including hippocampus, PRC, ERC, and PHC, were defined probabilistically in MNI space based on a database of manual MTL segmentations. (*C*) There was lower similarity for predictable versus unpredictable triads in PRC and ERC. Error bars indicate ±1 SEM. $^{**}P < 0.01$, $^{***}P < 0.001$.

observed in the hippocampus ($P = 0.32$) or PHC ($P = 0.63$). Critically, the predictable and unpredictable conditions were nearly identical (stimulus familiarity, action frequency, stimulus-transition frequency, etc.), with one key exception: on unpredictable trials, actions were not informative about how the cue would transition into the outcome. Triad-similarity effects were numerically stronger in the right hemisphere than in the left hemisphere for both PRC and ERC (Supplementary Fig. S5), but these differences were not reliable when the interaction between condition and hemisphere was tested ($Ps > 0.45$). Thus, PRC and ERC showed evidence of action-based object learning, representing triads as less similar to one another when their stimuli were linked by predictive actions.

Representational Space

Having obtained an overall effect of predictable triads being more distinguishable from each other in PRC and ERC, we can now examine the mechanism for why they were more distinguishable within these ROIs. Specifically, the finding of lower similarity between predictable triads in PRC and ERC can be explained by one of 2 types of representational change at the level of individual cue–action–outcome sequences that make up the triads (e.g., A1-left-B1 and A1-right-C1). According to a *merging* hypothesis (Fig. 4B), lower overall similarity between triads resulted from higher similarity between sequences within the same triad. That is, lower similarity between $A_1B_1C_1$ and $A_2B_2C_2$ may have resulted from higher similarity between A_1B_1 and A_1C_1, as well as between A_2B_2 and A_2C_2 (Fig. 4E). Alternatively, according a *differentiation* hypothesis (Fig. 4C), lower similarity between triads resulted from lower similarity between sequences from different triads. That is, lower similarity between $A_1B_1C_1$ and $A_2B_2C_2$ may have resulted from lower similarity between A_1B_1 and A_2C_2, and likewise between A_2B_2 and A_1C_1 (Fig. 4F). Importantly, although it is theoretically possible to obtain merging without differentiation and differentiation without merging (as displayed in Fig. 4), these types of representational change are not mutually exclusive and could together explain the triad-similarity difference between predictable and unpredictable triads.

We measured within-triad similarity as an index of merging (Fig. 5A), and across-triad similarity as an index of differentiation (Fig. 5C). In order to evaluate evidence for each hypothesis, we compared between predictable and unpredictable conditions separately for each measure. Note that the interaction between condition and measure is not informative since either increased within-triad similarity (merging) or decreased across-triad similarity (differentiation) could likewise lead to greater within-triad than across-triad similarity for the predictable condition but not the unpredictable condition. For both measures, we always compared across left and right button presses (i.e., the diagonal lines in Fig. 4). Comparing across left and right button presses was necessary by definition for measuring the within-triad similarity of predictable triads (since the 2 sequences had opposite responses). We wanted to equate this for the other analyses too (within-triad similarity of unpredictable triads and the across-triad similarity of all triads), to prevent a confound in which the response could be the same in some conditions but not others. All possible cue–action–outcome sequences were included in calculating the within-triad and across-triad similarity of both predictable and unpredictable triads.

Consistent with the merging hypothesis, we observed greater within-triad similarity between cue–action–outcome sequences

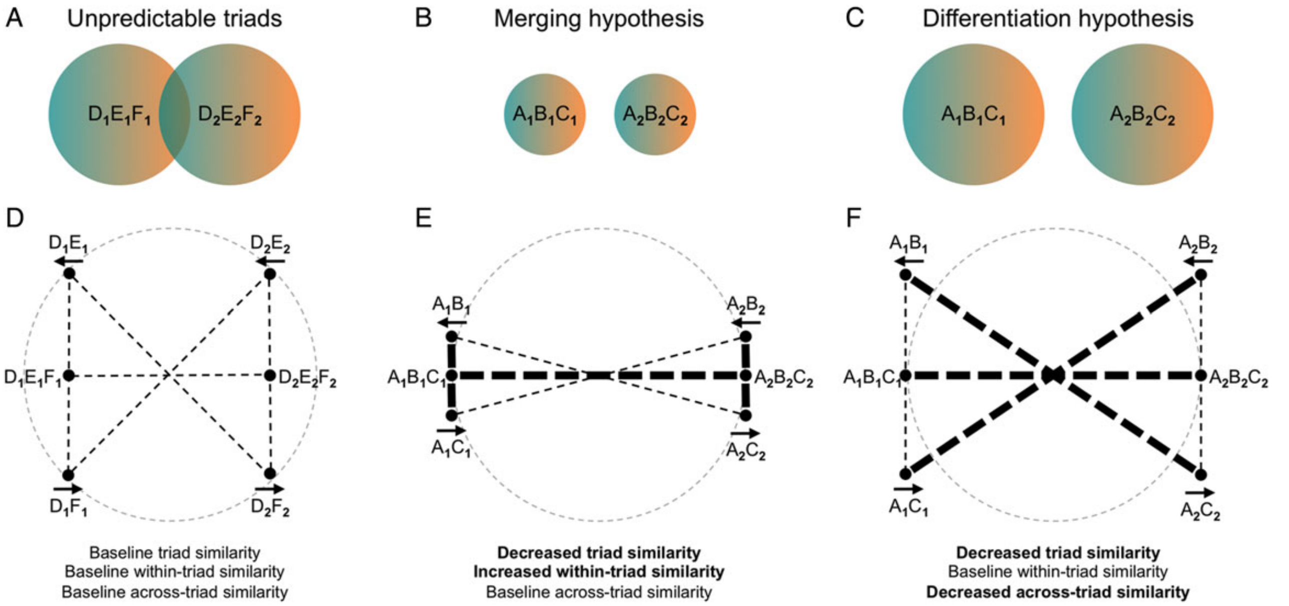
A
Unpredictable triads
D1E1F1
D2E2F2
B
Merging hypothesis
A1B1C1
A2B2C2
C
Differentiation hypothesis
A1B1C1
A2B2C2
D
D1E1
D2E2
D1E1F1
D2E2F2
D1F1
D2F2
Baseline triad similarity
Baseline within-triad similarity
Baseline across-triad similarity
E
A1B1
A2B2
A1B1C1
A2B2C2
A1C1
A2C2
Decreased triad similarity
Increased within-triad similarity
Baseline across-triad similarity
F
A1B1
A2B2
A1B1C1
A2B2C2
A1C1
A2C2
Decreased triad similarity
Baseline within-triad similarity
Decreased across-triad similarity

Figure 4. Merging versus differentiation. (*A*) A set-theoretic (Tversky and Gati 1982) intuition for merging and differentiation can be gained by visualizing the stimulus triads as overlapping distributions in representational space. (*B*) Lower triad similarity between predictable triads suggests that the underlying representations are less overlapping. This reduced overlap could occur if predictable triad representations became more tightly bound internally such that the size of each representation was smaller (merging). (*C*) Reduced overlap could also occur without any within-triad changes in representation, if the centers of the representations moved further apart from one another (differentiation). (*D*) A geometric (Shepard 1962) example of the alternative hypotheses can be gained by considering a representational space that is in the shape of a unit circle. Triad similarity can be decomposed into within-triad and across-triad similarity between the cue–action–outcome sequences that make up the triads. Similarity between the representations for unpredictable triads serves as a baseline to compare with similarity between representations for predictable triads. Non-bolded lines in (*D*–*F*) indicate this baseline similarity, while bolded lines in panels E and F indicate off-baseline similarity. (*E*) Merging hypothesis: lower similarity between triads reflects greater similarity between cue–action–outcome sequences within the same triad. (*F*) Differentiation hypothesis: lower triad similarity reflects lower similarity between cue–action–outcome sequences across different objects.

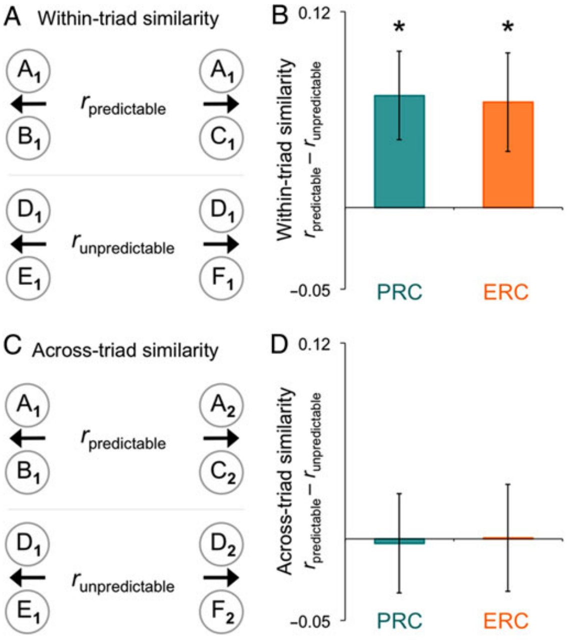

Figure 5. Within-triad versus across-triad similarity. (A) Merging predicts greater similarity of the cue–action–outcome sequences from within the same predictable triad, relative to an unpredictable triad. (B) This was found in both PRC and ERC. (C) Differentiation predicts reduced similarity of cue–action–outcome sequences across different predictable triads, relative to unpredictable triads. (D) Neither PRC nor ERC exhibited this effect. Error bars indicate ±1 SEM. *$P < 0.05$.

for predictable triads than for unpredictable triads (Fig. 5B), in both PRC ($t_{(23)} = 2.67$, $P = 0.01$) and ERC ($t_{(23)} = 2.15$, $P = 0.04$). Inconsistent with the differentiation hypothesis, there was no difference between predictable and unpredictable triads in across-triad similarity of cue–action–outcome sequences (Fig. 5D), in either PRC ($P = 0.86$) or ERC ($P = 0.95$). Thus, for a given cue, actions predictive of outcomes drew these cue–action–outcome sequences closer together in representational space.

For within-triad similarity, there was no interaction between condition and hemisphere in either PRC or ERC ($Ps > 0.28$). For across-triad similarity, there was a reliable interaction between condition and hemisphere in PRC ($P = 0.03$), but the simple effect of condition was not reliable in either hemisphere on its own ($Ps > 0.17$); there was no interaction in across-triad similarity between hemispheres of ERC ($P > 0.99$). To control for multiple comparisons, primary analyses of within-triad and across-triad similarity were restricted to PRC and ERC, the only MTL subfields that showed a triad-similarity difference between conditions. For completeness, however, we also examined whether the merging effect was selective to these regions (Supplementary Fig. S6). Indeed, within-triad similarity did not reliably differ between conditions in either the hippocampus ($P = 0.14$) or PHC ($P = 0.65$).

Subjective Predictability

We defined predictable versus unpredictable conditions based on the objective probabilities of the cue–action–outcome sequences. However, in addition to this, participants may have had a subjective impression of how predictable the sequences were on particular trials. Thus, it is unclear to what extent our effects mirror the actual statistics of the input or participants' conscious experience of the input. These 2 possibilities cannot, in fact, be distinguished for the predictable triads, since participants were required to reach a test criterion at which they were subjectively confident about the strong objective probabilities. However, the unpredictable triads provide such an opportunity: They were objectively unpredictable, and thus any behavior indicating that participants found them predictable can be interpreted as subjective.

We used test performance as a measure of whether participants found individual unpredictable triads to be predictable. We quantified how consistently participants mapped each outcome onto specific cue/action combinations across the 3 behavioral tests. A subjective association was inferred when an outcome was chosen with 100% consistency across all test sessions on both days of the experiment. For example, if a participant always chose E as the outcome for D with a left response and F as the outcome for D with a right response, then we categorized this triad as subjectively predictable (likewise, any inconsistency across sessions was categorized as subjective unpredictability). Participants varied in consistency: 3 were consistent for both of their unpredictable triads, 8 were consistent for one unpredictable triad but inconsistent for the other, and 13 were inconsistent for both unpredictable triads.

Is subjective predictability, defined this way, sufficient to induce merging in MTL? Or does action-based learning depend expressly on the objective predictability of the outcomes? Because the amount of merging based on objective predictability was statistically indistinguishable in PRC and ERC ($P = 0.94$), we pooled these voxels into a single MTL ROI in order to address these follow-up questions with greater power and fewer comparisons. To test the effect of subjective experience on merging, we measured the pattern similarity of the 2 cue–action–outcome sequences within 3 types of triads (Fig. 6A): predictable triads with consistent test responses, unpredictable triads with consistent test responses, and unpredictable triads with inconsistent test responses. [Despite the uniformly consistent responses for predictable triads in the criterion test, there were 4 predictable triads with inconsistent responses in later tests across all participants (out of a total of 48 predictable triads). This was an insufficient number to analyze separately and thus they were excluded from this analysis (but included in earlier analyses).] Note that, for unpredictable-consistent and unpredictable-inconsistent triads, this analysis focused on trials in which the outcome matched the preferred response mapping at test (see Materials and Methods).

Because consistency varied across participants, and not every participant had both consistent and inconsistent unpredictable triads, we examined the effect of subjective predictability by pooling normalized data across participants (Fig. 6). We then used subject-level bootstrap resampling (Efron and Tibshirani 1986) to test the random-effects significance of the relationship between MTL within-triad similarity and test consistency (Supplementary Fig. S7). As expected based on the separate analysis of PRC and ERC ROIs, predictable-consistent triads produced more within-triad pattern similarity in MTL than unpredictable-inconsistent triads ($P = 0.001$). Interestingly, among unpredictable triads, pattern similarity was reliably greater for consistent versus inconsistent triads ($P = 0.01$). Indeed, there was no difference in within-triad similarity between predictable-consistent and unpredictable-consistent triads ($P = 0.99$).

These merging effects could be interpreted as reflecting the expectations induced by subjectively predictive cues and actions and/or the fulfillment of these expectations when the anticipated outcome appeared. To evaluate this latter possibility, we compared unpredictable-consistent trials in which the expected outcome appeared (as used in the merging analysis above) to

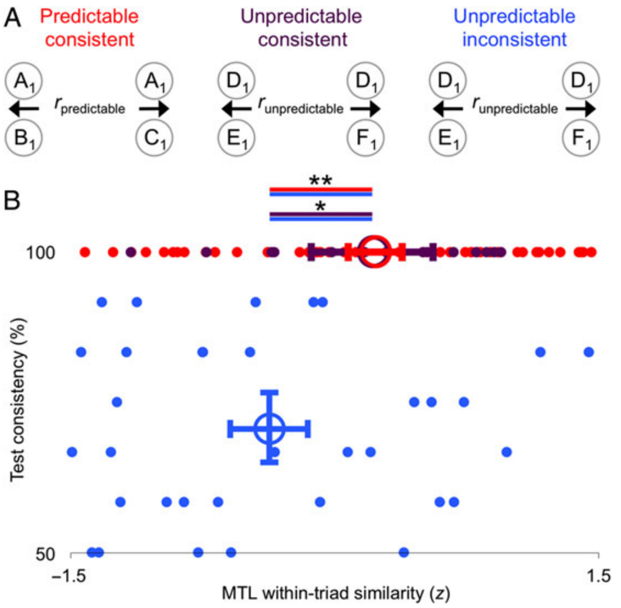

Figure 6. Test consistency. (A) Within-triad similarity (an index of merging) was measured separately for each triad type within a combined PRC/ERC ROI. (B) Merging was greater for predictable-consistent triads and unpredictable-consistent triads, relative to unpredictable-inconsistent triads. Open circles show the average pattern similarity and consistency for each triad type, while error bars indicate 95% confidence intervals for each average, derived from bootstrap resampling over participants. *$P < 0.05$, **$P < 0.01$.

unpredictable-consistent trials in which the other, unexpected outcome occurred. The same expectation was present in both cases—the same cues and actions were involved—but differed in terms of whether this expectation was fulfilled or violated. The match between expectation and outcome played a role in merging, with marginally greater within-triad similarity when the expected versus unexpected outcome appeared ($P = 0.07$; Supplementary Fig. S7D). This analysis does not rule out a contribution of the expectation itself, but suggests that its fulfillment was also important.

Consequences for the Visual System

How does action-based learning in the MTL interact with perceptual processing in visual cortex? Prior studies of object learning and recognition have focused on posterior, object-selective visual areas such as LO (Grill-Spector et al. 2001). These studies consistently report attenuated activity in LO for repeated versus novel stimuli (Schacter et al. 2007), with growing evidence for increased repetition attenuation when repeated stimuli are expected versus unexpected (Summerfield et al. 2008; Larsson and Smith 2012); these latter findings have been interpreted as reflecting reduced prediction error (Ewbank and Henson 2012).

To provide a baseline measure of repetition attenuation, a subset of trials in the scan task included a predictable or unpredictable cue stimulus followed by a novel-outcome stimulus that was not seen at all during training (Supplementary Fig. S1). In comparing predictable and unpredictable triads (with familiar outcomes) to these trials, we found a reliable reduction in activation for both LO ($F_{1,23} = 57.88$, $P < 0.001$) and the combined PRC/ERC ROI ($F_{1,23} = 7.72$, $P = 0.01$). Mean activity did not reliably differ between predictable and unpredictable triads in either MTL or LO ($Ps > 0.7$), and the interaction between predictable and unpredictable triads and the novel-outcome stimuli did not approach significance for either region ($Ps > 0.3$). Based on the overall effect of repetition attenuation for familiar outcome stimuli, we used reduced univariate activation in LO as a proxy for facilitated perceptual processing.

Learning the structure of predictable triads could allow participants to form an expectation about which outcome should appear given a cue and action. We reasoned that such expectations might influence how outcomes are processed in LO. Specifically, we tested for a trial-by-trial relationship between the amount of within-triad similarity in the MTL ROI (PRC/ERC) and the amount of repetition attenuation in LO (Supplementary Fig. S8). We hypothesized that the more that a full triad representation was reactivated in MTL from the cue–action–outcome sequence on a given trial, the more strongly that the learned outcome could be predicted, and the lower the prediction error (and hence activity) in LO when the outcome appeared in that sequence. We expected this negative relationship for predictable-consistent triads but not for unpredictable-inconsistent triads. If subjective predictability was again sufficient, we should also obtain a negative relationship for unpredictable-consistent triads.

For predictable-consistent triads, reactivation in MTL negatively predicted LO activity across trials ($P = 0.001$; Fig. 7A). This effect specifically tracked objective probabilities as it was not found for unpredictable-consistent ($P = 0.58$; Fig. 7B), nor for unpredictable-inconsistent triads ($P = 0.26$; Fig. 7C). The comparisons of predictable-consistent triads to unpredictable-consistent triads ($P = 0.04$) and to unpredictable-inconsistent triads ($P = 0.03$) were reliable; unpredictable-consistent triads and unpredictable-inconsistent triads did not differ ($P = 0.87$). Furthermore, when the relationship between MTL triad reactivation and attenuation was examined over the whole brain, voxels that were identified as stimulus selective in the localizer were more likely to exhibit this relationship in the predictable-consistent than unpredictable-consistent condition ($\chi^2(1, N = 24) = 15.35$, $P < 0.001$); no voxels exhibited a reliable relationship for unpredictable-inconsistent (Supplementary Fig. S9).

Although merging was found for both predictable-consistent and unpredictable-consistent triads when MTL was considered in isolation, examining MTL–LO interactions revealed a dissociation, with only predictable-consistent exhibiting an effect. In other words, the conscious experience of predictive action was sufficient to influence MTL representations themselves, but not to enable these representations to impact processing in stimulus-selective visual cortex.

Discussion

We employed a novel paradigm in which actions induced predictable versus unpredictable transitions between stimuli. Subregions of MTL cortex — the PRC and ERC in particular — came to represent triads of stimuli that were linked by predictive actions as less similar to other triads. This increased distance between triads was driven by representational merging, whereby the alternative cue–action–outcome sequences from the same triad produced more similar activity patterns. Among triads for which the outcomes could not be predicted from actions objectively, similar merging occurred in MTL when participants subjectively felt that their actions predicted the outcomes. Furthermore, the reactivation of triad representations in MTL was linked to reduced activity in stimulus-selective visual cortex, consistent with the possibility that predictive actions create expectations that obviate sensory processing. This interaction was restricted to objectively predictable stimulus triads. We interpret these findings as evidence of a new form of action-based learning

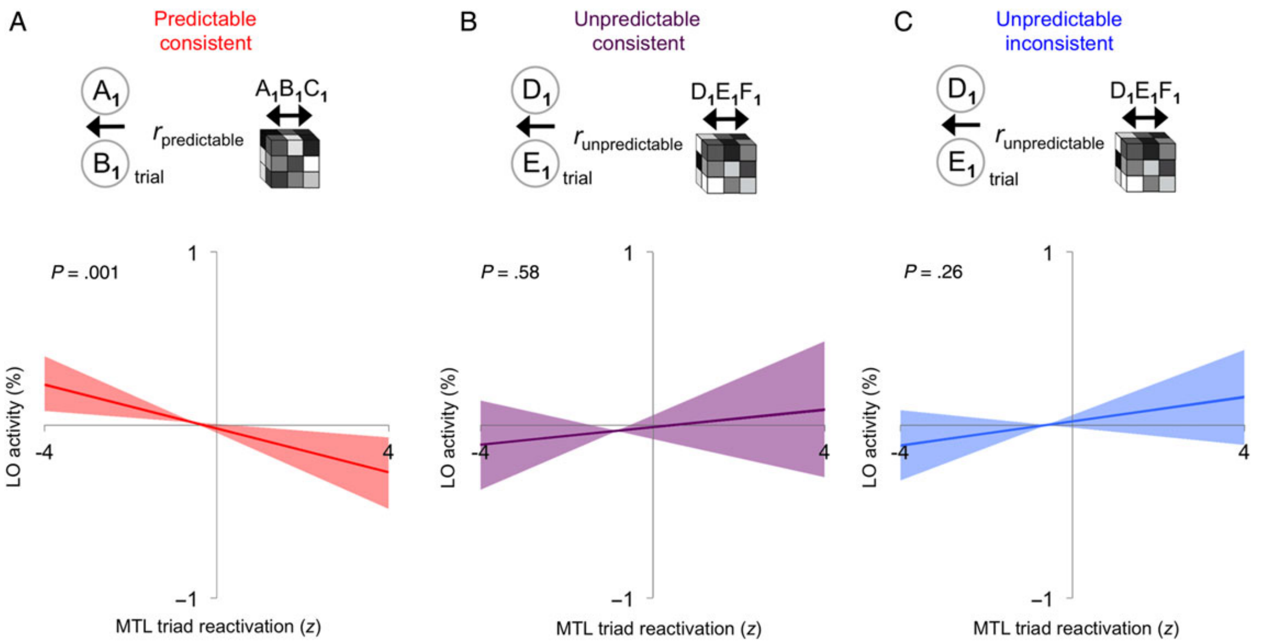

Figure 7. MTL–LO interaction. (*A*) Triad reactivation in PRC/ERC was associated with lower LO activity across trials generated from predictable-consistent triads, but not from either (*B*) unpredictable-consistent triads or (C) unpredictable-inconsistent triads. Error bands indicate 95% confidence intervals of the linear trends, derived from bootstrap resampling.

in the MTL, whereby actions help create structured representations about the different possible states of objects.

Because stimuli were randomly assigned to predictable and unpredictable triads, the overlap of static visual features among stimuli was arbitrary. This is distinct from other types of object learning, such as forming invariant representations across subtle changes in viewpoint, where the stimuli share many if not most of the same features (Biederman and Gerhardstein 1993; DiCarlo et al. 2012). The lack of feature overlap implicates the MTL, given its ability to bind arbitrary stimuli such as in statistical learning (Schapiro et al. 2012) and episodic encoding (Davachi 2006; Eichenbaum et al. 2007). Within the MTL, the effects we observed were restricted to the PRC and ERC rather than the hippocampus. This may reflect the fact that the triads were learned over multiple days and thus had the opportunity to be consolidated (Erickson and Desimone 1999; Norman and O'Reilly 2003).

By isolating merging without differentiation as the representational change underlying the formation of object identities for predictable stimulus triads, we strictly constrain the hypothesis space of potential cognitive and neural mechanisms that could underlie action-based learning. Specifically, merging suggests the formation of schematic knowledge structures for predictable triads. Although the analyses used for examining action-based learning, such as measuring triad reactivation for each trial, may specifically suggest a prototype-based model of category learning (Reed 1972; Rosch 1978), the observed merging without differentiation is also consistent with the formation of knowledge structures through exemplar-based models (Medin and Schaffer 1978; Hintzman 1986). In contrast, merging contradicts the possibility that the triad similarity effect was due to interference for unpredictable triads (Kirwan and Stark 2007; Axmacher et al. 2009; Watson and Lee 2013). Such interference would cause greater within-triad similarity for unpredictable triads than for predictable triads, since the multivariate pattern for all trials (regardless of action) of an unpredictable triad would reflect reactivation of both outcomes. Likewise, greater within-triad similarity for unpredictable triads would be expected if only one outcome was predicted per trial, but which outcome was predicted varied across trials. In this case, predictions for both outcomes would be evident in the average pattern across trials.

Merging described here also relates to more basic neural mechanisms of pattern completion and pattern separation (Marr 1971; O'Reilly and Rudy 2001). For instance, the observed effects are consistent with pattern completion for predictable triads, in which individual cue–action–outcome sequences reactivate the complete representation of a stimulus triad, including the other possible sequence for the cue. Insofar as such retrieval causes cortical reinstatement, pattern completion could provide the mechanism by which action-based learning facilitates visual processing. It is important to note, however, that these processes are often ascribed to the hippocampus (Leutgeb and Leutgeb 2007; Bakker et al. 2008). Although triad-level representations in hippocampus did not differ between conditions, hippocampal processes of pattern completion and pattern separation may in turn support the higher-level object learning observed here in ERC and PRC.

In prior studies of arbitrary associative learning in the MTL, the stimulus sequence itself contained all of the needed information about how stimuli should be associated (Naya and Suzuki 2011; Schapiro et al. 2012). That is, to-be-associated stimuli could be identified based on having higher transition probabilities with each other than with other stimuli. However, this was not the case in the current study when comparing predictable with unpredictable triads: in both conditions, the cue was exactly 50% likely to be followed by each of the outcomes. Although arrows prompted actions in the scanner for counterbalancing purposes, this was critically not the case during exploratory training that was performed until learning reached criterion. Throughout exploratory training, the *only* information that distinguished conditions was that actions made the outcomes perfectly deterministic for predictable triads but did not add any predictive value to the outcomes for unpredictable triads. The merging observed in the scanner for predictable triads may suggest that actions become an integral part of the object representation (e.g., that the cue and outcomes are bound to the actions in the MTL). However, it is also possible that this finding reflects stimulus–stimulus learning that was initially facilitated by actions (e.g., predictive actions were a sign that 2 stimuli belonged together). One approach for disentangling these interpretations may be to present a cue, elicit an action, and then show a blank screen. Any effect of the action on outcome reactivation in MTL or LO must reflect the fact that the action is part of the representation. Alternatively, one could present the cue and outcome in the scanner without an action and see whether the predictable versus unpredictable effects persist. If so, this would suggest that actions facilitate stimulus–stimulus learning but are not required for its expression once at asymptote.

We view the MTL as central not only for learning the structure of the triads, but also for deploying this knowledge subsequently via the retrieval of learned associations in the service of prediction. The observed interaction between MTL triad reactivation and LO activity for predictable triads is consistent with this possibility, and it extends prior findings about the role of MTL modulating visual cortex (Bosch et al. 2014) to include prediction from action. Such action-based perceptual prediction can be described within a hierarchical Bayesian framework in which many sources of expectation converge to "explain away" sensory evidence and minimize prediction error (Rao et al. 1999; Friston 2005). Previously considered sources of visual expectation include recent stimuli (Summerfield et al. 2008), auditory cues (Kok et al. 2012), temporal context (Turk-Browne et al. 2012), and actions alone (Cardoso-Leite et al. 2010). In the current case, predicting a forthcoming stimulus involved an initial stimulus and an action, which were together coupled in the MTL as part of a multistate object representation.

MTL–LO interactions were evident for objectively predictable stimulus triads, and not for subjectively predictable (i.e., unpredictable-consistent) triads. There are several possible explanations for this dissociation. For instance, the dissociation may simply reflect the differences in sensitivity between pattern analysis (of MTL) and univariate analysis (of visual cortex) for detecting cognitive states (Norman et al. 2006; Davis et al. 2014). Alternatively, the dissociation may suggest distinct mechanisms for the formation of object representations in MTL and the effect of these representations on visual cortex. Furthermore, although a subjective association was inferred when an outcome was chosen with 100% consistency across all test sessions on both days of the experiment, it is possible that these were doubtful associations for at least a subset of participants. For instance, a participant may have designated a particular set of left/right response mappings for one of the unpredictable triads on the first behavioral test. Then, in the 2 subsequent tests, the participant may remember previously indicated left/right mappings, and continue to indicate this initial response mapping for the triad despite increased uncertainty about the validity of the preferred mapping. In this case, the participant would have maintained the knowledge that a specific outcome was previously associated with the cue and a left or right action. However, the participant's perceptual expectation that the preferred outcome would actually appear for a particular trial would be relatively weak. A more direct measurement of confidence, such as confidence ratings, could potentially reveal effects of subjective predictability on visual cortex that were not observed here.

Although we considered very simple objects and a limited repertoire of actions in our study, such action-based learning may scale up to more complex objects and actions. Do associations between complex actions and real-world objects experienced over a lifetime lead to similar learned representations as those observed here? Considering that semantic representations of real-world objects may include the specific action affordances of those objects (Gibson 1986; Cisek and Kalaska 2010; Valyear et al. 2012), future studies may test whether preexisting semantic knowledge can be leveraged to generate predictions about object states. Most real-world actions would be difficult to perform in the scanner, but studies of action simulation and mirror neurons (Calvo-Merino et al. 2005; Rizzolatti et al. 2014) suggest that effects of action-based learning could potentially be observed in an experimental setting in which motor responses are withheld.

Our mind normally infers that 2 stimuli belong to the same object based on perceptual cues such as feature similarity and spatiotemporal continuity (Pylyshyn 1989; Kahneman et al. 1992; Yi et al. 2008; Flombaum et al. 2009). The present findings suggest that beyond these factors, learning about which stimuli follow each other and how our actions control this transition may be important for object perception.

Supplementary Material

Supplementary Material can be found at http://www.cercor.oxfordjournals.org/.

Funding

This work was funded by National Institutes of Health R01-EY021755 to N.B.T.-B. and National Institutes of Health F32-EY023162 to N.C.H.

Notes

The authors thank Aaron Bornstein and Janice Chen for helpful comments on an earlier draft. *Conflict of Interest*: None declared.

References

Andersson JL, Jenkinson M, Smith S. 2007. Non-linear registration, aka spatial normalisation. Technical Report TR07JA2, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, Department of Clinical Neurology. Oxford University, Oxford, UK.

Axmacher N, Haupt S, Cohen MX, Elger CE, Fell J. 2009. Interference of working memory load with long-term memory formation. Eur J Neurosci. 29:1501–1513.

Bakker A, Kirwan CB, Miller M, Stark CE. 2008. Pattern separation in the human hippocampal CA3 and dentate gyrus. Science. 319:1640–1642.

Biederman I, Gerhardstein PC. 1993. Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. J Exp Psychol Hum Percept Perform. 19:1162.

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

Calvo-Merino B, Glaser DE, Grèzes J, Passingham RE, Haggard P. 2005. Action observation and acquired motor skills: An fMRI study with expert dancers. Cereb Cortex. 15:1243–1249.

Cardoso-Leite P, Mamassian P, Schütz-Bosbach S, Waszak F. 2010. A new look at sensory attenuation action: effect anticipation affects sensitivity, not response bias. Psychol Sci. 21:1740–1745.

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

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

Cisek P, Kalaska JF. 2010. Neural mechanisms for interacting with a world full of action choices. Annu Rev Neurosci. 33: 269–298.

Cohen NJ, Eichenbaum H. 1993. Memory, amnesia, and the hippocampal System. Cambridge, MA: MIT press.

Cox RW. 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res. 29:162–173.

Dale AM. 1999. Optimal experimental design for event-related fMRI. Hum Brain Mapp. 8:109–114.

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

Davis T, LaRocque KF, Mumford JA, Norman KA, Wagner AD, Poldrack RA. 2014. What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. Neuroimage. 97:271–283.

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

Duvernoy HM. 2005. The human hippocampus: functional anatomy, vascularization and serial sections with MRI. NY: Springer.

Efron B, Tibshirani R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci. 30:54–75.

Eichenbaum H, Yonelinas AR, Ranganath C. 2007. The medial temporal lobe and recognition memory. Annu Rev Neurosci. 30:123–152.

Erickson CA, Desimone R. 1999. Responses of macaque perirhinal neurons during and after visual stimulus association learning. J Neurosci. 19:10404–10416.

Ewbank MP, Henson RN. 2012. Explaining away repetition effects via predictive coding. Cogn Neurosci. 3:239–240.

Flombaum JI, Scholl BJ, Santos LR. 2009. Spatiotemporal priority as a fundamental principle of object persistence. In: Hood B, Santos L, editors. The origins of object knowledge. London: Oxford University Press. p.135–164

Friston K. 2005. A theory of cortical responses. Phil Trans R Soc Lond B. 360:815–836.

Gibson JJ. 1986. The ecological approach to visual perception. Hillsdale, NJ: Lawrence Erlbaum.

Greve DN, Fischl B. 2009. Accurate and robust brain image alignment using boundary-based registration. Neuroimage. 48:63–72.

Grill-Spector K, Kourtzi Z, Kanwisher N. 2001. The lateral occipital complex and its role in object recognition. Vision Res. 41:1409–1422.

Henson RN, Gagnepain P. 2010. Predictive, interactive multiple memory systems. Hippocampus. 20:1315–1326.

Hindy NC, Altmann GTM, Kalenik E, Thompson-Schill SL. 2012. The effect of object state-changes on event processing: do objects compete with themselves? J Neurosci. 32:5795–5803.

Hindy NC, Solomon SH, Altmann GTM, Thompson-Schill SL. 2015. A cortical network for the encoding of object change. Cereb Cortex. 25:884–894.

Hintzman DL. 1986. "Schema abstraction" in a multiple-trace memory model. Psychol Rev. 93:411–428.

Kahneman D, Treisman A, Gibbs BJ. 1992. The reviewing of object files: object-specific integration of information. Cogn Psychol. 24:175–219.

Kim G, Lewis-Peacock JA, Norman KA, Turk-Browne NB. 2014. Pruning of memories by context-based prediction error. Proc Natl Acad Sci USA. 201319438.

Kirwan CB, Stark CEL. 2007. Overcoming interference: an fMRI investigation of pattern separation in the medial temporal lobe. Learn Mem. 14:625–633.

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

Kourtzi Z, Kanwisher N. 2000. Cortical regions involved in perceiving object shape. J Neurosci. 20:3310–3318.

Larsson J, Smith AT. 2012. fMRI repetition suppression: Neuronal adaptation or stimulus expectation? Cereb Cortex. 22:567–576.

Leutgeb S, Leutgeb JK. 2007. Pattern separation, pattern completion, and new neuronal codes within a continuous CA3 map. Learn Mem. 14:745–757.

Marr D. 1971. Simple memory: a theory for archicortex. Phil Trans R Soc B. 262:23–81.

Medin DL, Schaffer MM. 1978. Context theory of classification learning. Psychol Rev. 85:207–238.

Miyashita Y. 1993. Inferior temporal cortex: where visual perception meets memory. Annu Rev Neurosci. 16:245–263.

Miyashita Y. 1988. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. Nature. 335:817–820.

Naya Y, Suzuki WA. 2011. Integrating what and when across the primate medial temporal lobe. Science. 333:773–776.

Niv Y, Schoenbaum G. 2008. Dialogues on prediction errors. Trends Cogn Sci. 12:265–272.

Norman KA, O'Reilly RC. 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol Rev. 110:611.

Norman KA, Polyn SM, Detre GJ, Haxby JV. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci. 10:424–430.

Op de Beeck HP. 2010. Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? Neuroimage. 49:1943–1948.

O'Reilly RC, Rudy JW. 2001. Conjunctive representations in learning and memory: principles of cortical and hippocampal function. Psychol Rev. 108:311.

Poldrack RA, Clark J, Pare-Blagoev EJ, Shohamy D, Moyano JC, Myers C, Gluck MA. 2001. Interactive memory systems in the human brain. Nature. 414:546–550.

Pylyshyn Z. 1989. The role of location indexes in spatial perception: a sketch of the FINST spatial-index model. Cognition. 32:65–97.

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

Reed SK. 1972. Pattern recognition and categorization. Cogn Psychol. 3:382–407.

Rizzolatti G, Cattaneo L, Fabbri-Destro M, Rozzi S. 2014. Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. Physiol Rev. 94:655–706.

Rosch E. 1978. Principles of categorization. In: Cognition and categorization. Hillsdale, NJ: Erlbaum. p. 27–48.

Sadeh T, Shohamy D, Levy DR, Reggev N, Maril A. 2010. Cooperation between the hippocampus and the striatum during episodic encoding. J Cogn Neurosci. 23:1597–1608.

Sayres R, Grill-Spector K. 2008. Relating retinotopic and object-selective responses in human lateral occipital cortex. J Neurophysiol. 100:249–267.

Schacter DL, Wig GS, Stevens WD. 2007. Reductions in cortical activity during priming. Curr Opin Neurobiol. 17:171–176.

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

Shepard RN. 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function. I Psychometrika. 27:125–140.

Shohamy D, Turk-Browne NB. 2013. Mechanisms for widespread hippocampal involvement in cognition. J Exp Psychol Gen. 142:1159.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, et al. 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage Math Brain Imaging. 23(Suppl 1): S208–S219.

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

Turk-Browne NB, Simon MG, Sederberg PB. 2012. Scene representations in parahippocampal cortex depend on temporal context. J Neurosci. 32:7202–7207.

Tversky A, Gati I. 1982. Similarity, separability, and the triangle inequality. Psychol Rev. 89:123.

Valyear KF, Gallivan JP, McLean DA, Culham JC. 2012. fMRI repetition suppression for familiar but not arbitrary actions with tools. J Neurosci. 32:4247–4259.

Watson HC, Lee ACH. 2013. The perirhinal cortex and recognition memory interference. J Neurosci. 33:4192–4200.

Wimmer GE, Shohamy D. 2012. Preference by association: how memory mechanisms in the hippocampus bias decisions. Science. 338:270–273.

Wirth S, Yanike M, Frank LM, Smith AC, Brown EN, Suzuki WA. 2003. Single neurons in the monkey hippocampus and learning of new associations. Science. 300:1578.

Yi DJ, Turk-Browne NB, Flombaum JI, Kim MS, Scholl BJ, Chun MM. 2008. Spatiotemporal object continuity in human ventral visual cortex. Proc Natl Acad Sci USA. 105:8840.

Yin HH, Knowlton BJ. 2006. The role of the basal ganglia in habit formation. Nat Rev Neurosci. 7:464–476.

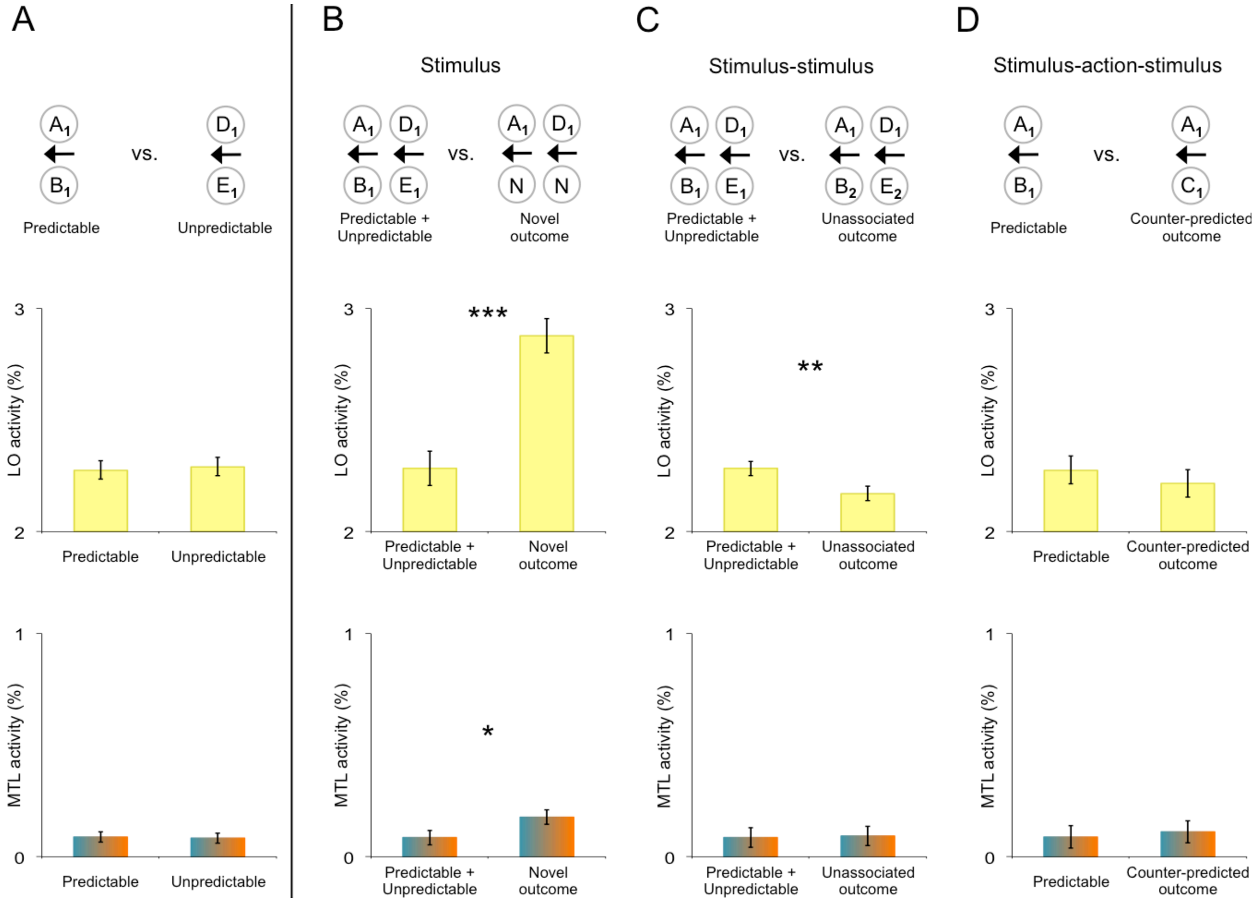

Figure S1. Repetition effects on univariate activity. The left-hand side of each comparison includes learned stimuli considered in the primary analyses in order to measure pattern similarity and cortical interactions. The right-hand side of each comparison includes additional stimuli presented during the scan task in order to provide baselines for measuring repetition-related univariate activity in each region. **(A)** To measure univariate effects of predictability, we compared the average level of activity elicited for predictable triads to the average level of activity elicited for unpredictable triads. Mean activity did not reliably differ between predictable and unpredictable triads in either region (Ps > .7). **(B)** To measure activity for stimulus repetition, we compared activity for predictable and unpredictable triads to activity for an additional subset of trials that included a novel outcome stimulus that was not seen at all during training. A reliable main effect of decreased activity was observed in both LO ($F(1, 23) = 57.88$, $P < .001$) and the PRC/ERC ROI for MTL ($F(1, 23) = 7.72$, $P = .01$), while the interaction between predictable and unpredictable triads did not approach significance for either region (Ps > .3). **(C)** To measure activity for repeated stimulus-stimulus transitions, we compared activity for predictable and unpredictable triads to activity for an additional subset of trials that included an unassociated outcome that had been paired with a different cue during training. A reliable main effect of increased activity was observed in LO ($F(1, 23) = 7.72$, $P = .01$), but not in MTL ($P = .65$), while the interaction between predictable and unpredictable triads did not approach significance for either region (Ps > .4). **(D)** To measure activity for repeated stimulus-action-stimulus transitions, we compared activity for predictable triads to activity for an additional subset of trials that included a counter-predicted outcome that was expected only after the alternative (left vs. right) action. Activity did not reliably differ between learned and novel stimulus-action-stimulus transitions in either region (Ps > .3). Error bars indicate ± 1 SEM of the difference between conditions. N = novel outcome, *$P < .05$, **$P < .01$, ***$P < .001$.

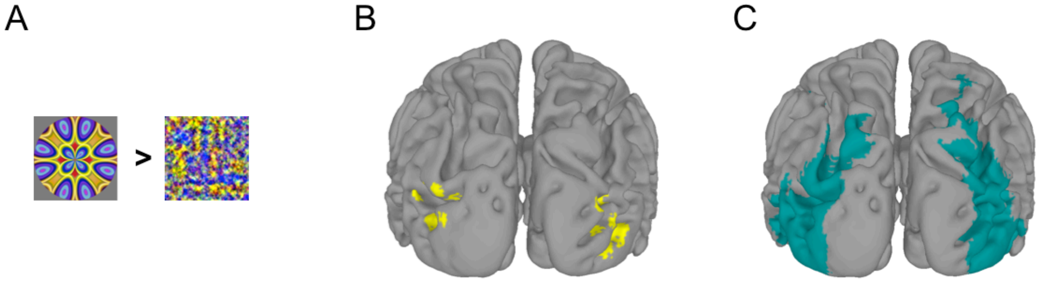

Figure S2. Stimulus-selective LO. The final scanning run was a localizer in which participants viewed blocked sequences of fractal patterns, alternating with fixation and with blocks of phase-scrambled images of the fractal patterns. **(A)** We used the contrast of fractal > scramble to identify stimulus-selective voxels. **(B)** LO ROIs (overlay of individual ROIs depicted) were defined in the left and right hemisphere for each participant as 3-mm spheres, each centered on the peak voxel in LO. **(C)** For voxelwise analyses (Supplementary Fig. S1), stimulus-selective cortex was identified across participants at a whole-brain corrected α of $P < .05$, based on a voxelwise α of $P < .001$ and a cluster-forming threshold of 25 voxels determined using 3dClustSim (Cox 1996; http://afni.nimh.nih.gov/pub/dist/doc/program_help/3dClustSim.html).

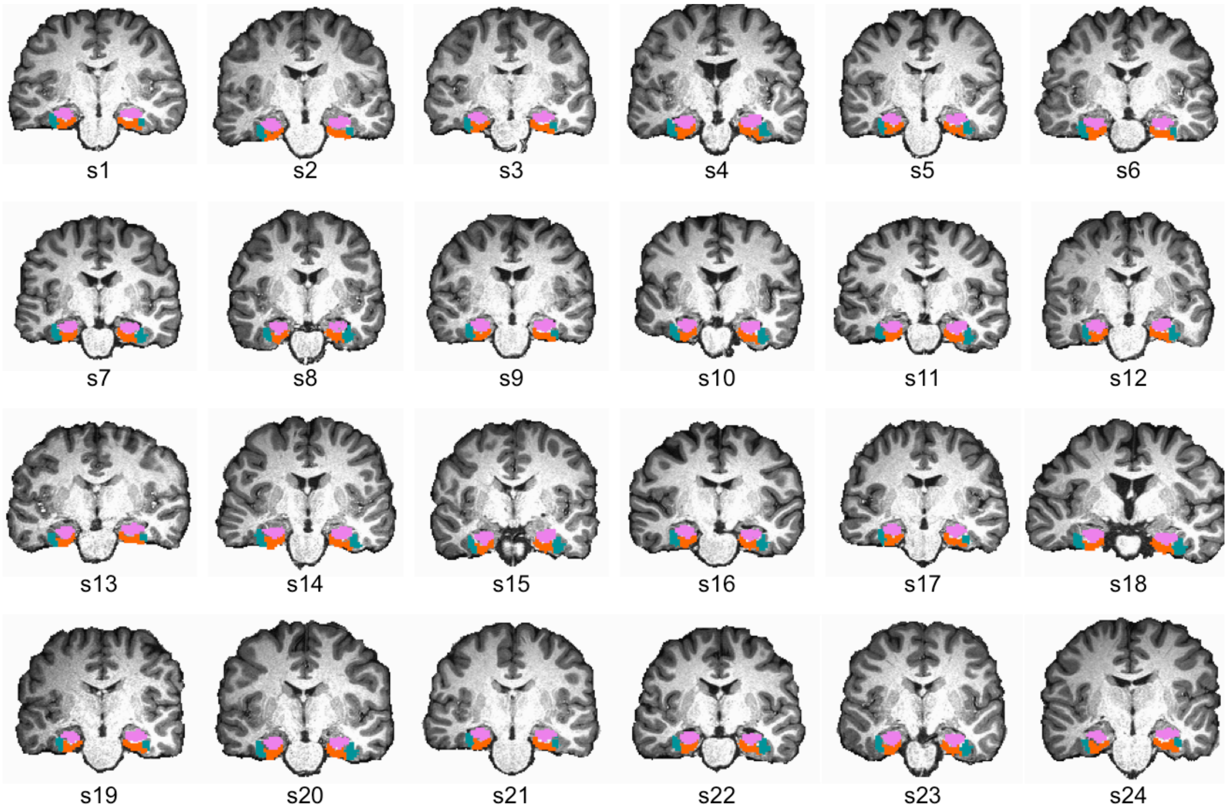

Figure S3. MTL ROIs for each participant in high-resolution anatomical space. MTL ROIs, defined probabilistically based on a database of manual segmentations, highly conformed to each participant's anatomy. Hippocampus (purple), PRC (green), and ERC (orange) are displayed as overlays on a high-resolution anatomical image for each participant. PHC is not displayed.

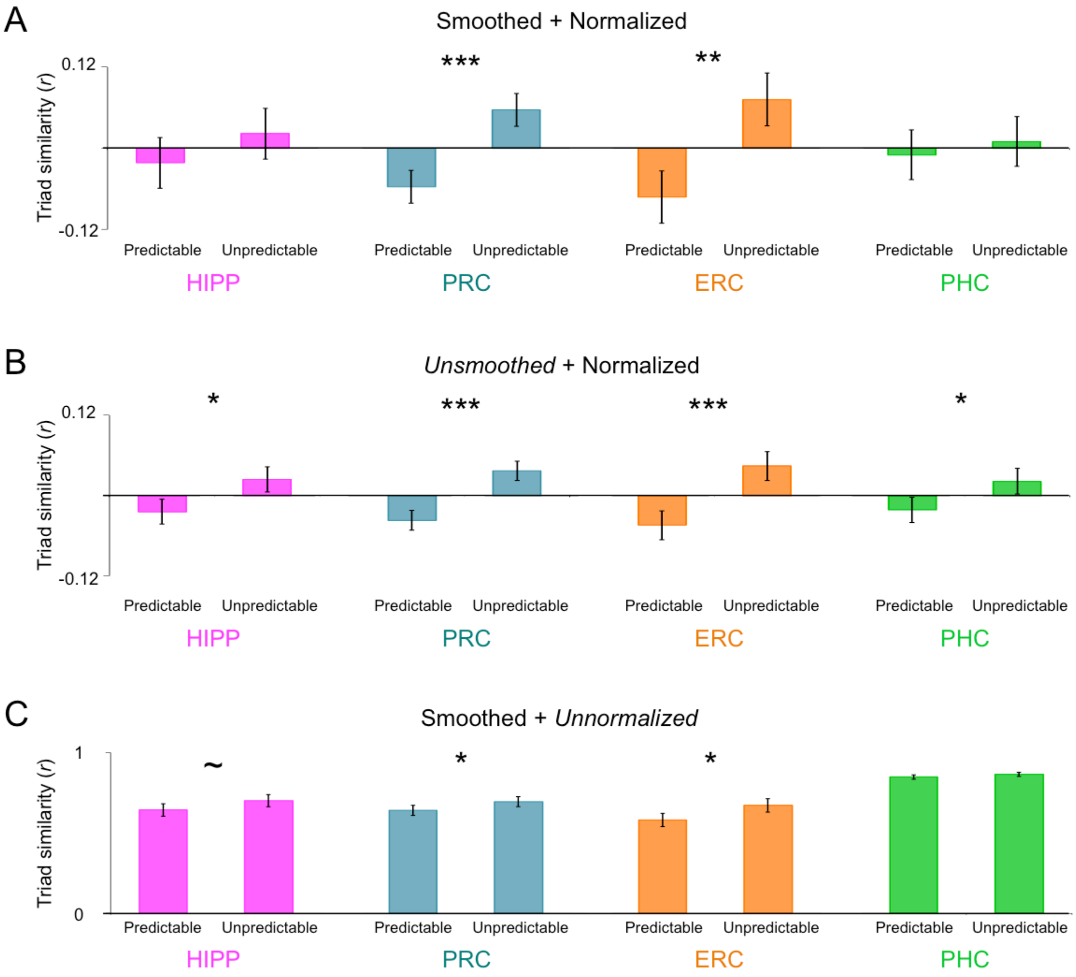

Figure S4. Effects of smoothing and normalization on triad similarity. (A) Data plotted in Fig 3C was spatially smoothed and normalized within-voxel across parameter estimates. We calculated the pattern similarity between the 2 predictable triads and between the 2 unpredictable triads in ROIs for hippocampus, PRC, ERC, and PHC. There was lower similarity for predictable vs. unpredictable triads in PRC ($t(23) = 4.87$, $P < .001$) and ERC ($t(23) = 3.75$, $P = .001$), but no difference in either hippocampus or PHC (*P*s > .32). **(B)** When parameter estimates were normalized but no spatial smoothing was applied, we observed reliable triad similarity differences between conditions in PRC ($t(23) = 5.07$, $P < .001$) and ERC ($t(23) = 4.01$, $P < .001$), and also in hippocampus ($t(23) = 2.45$, $P = .02$) and PHC ($t(23) = 2.20$, $P = .04$). **(C)** When patterns were spatially smoothed and parameter estimates were unnormalized, we observed a reliable difference between conditions in PRC ($t(23) = 2.46$, $P = .02$) and ERC ($t(23) = 2.79$, $P = .01$), along with a marginally reliable difference in hippocampus ($t(23) = 1.88$, $P = .07$) and no difference in PHC ($P = .24$). Error bars indicate ± 1 SEM. *$P < .05$, **$P < .01$, ***$P < .001$.

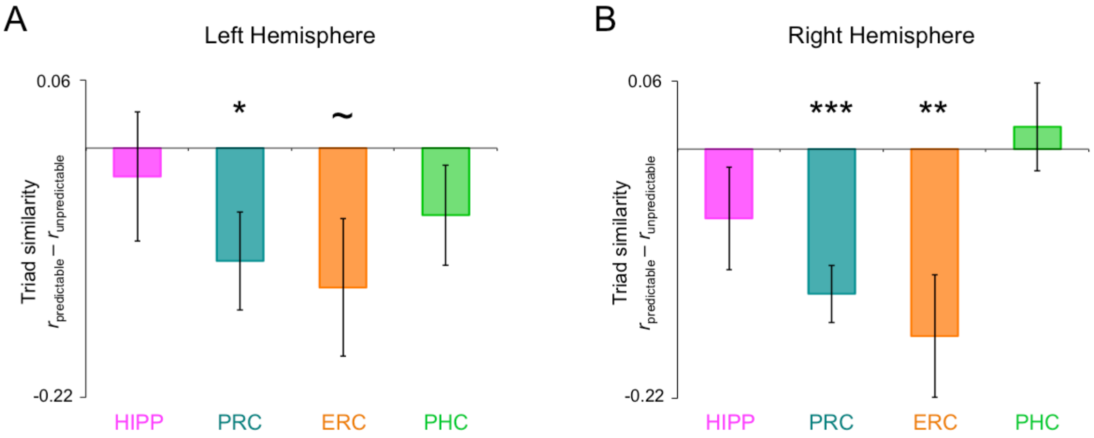

Figure S5. Triad similarity in each hemisphere. (A) Triad similarity in left PRC was reliably lower for predictable triads than for unpredictable triads ($t(23) = 2.22$, $P = .04$), and there was a marginally reliable difference between conditions in left ERC ($t(23) = 2.00$, $P = .06$). There was no difference between conditions in either left hippocampus or left PHC (*P*s > .25). **(B)** Triad similarity was reliably lower for predicable triads in both right PRC ($t(23) = 5.08$, $P < .001$) and right ERC ($t(23) = 3.19$, $P = .004$). There was no difference between conditions in either right hippocampus or right PHC (*P*s > .18). The interaction between condition and hemisphere was not reliable for any of the ROIs (*P*s > .20). Error bars indicate ± 1 SEM. ˜$P = .06$, *$P < .05$, **$P < .01$, ***$P < .001$**.**

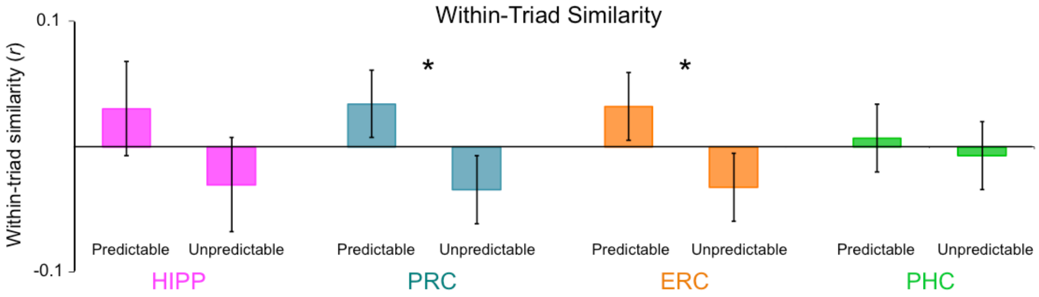

Figure S6. Within-triad similarity for each MTL ROI. Greater within-triad similarity for predictable triads than for unpredictable triads was found in both PRC (*t*(23) = 2.67, *P* = .01) and ERC (*t*(23) = 2.15, *P* = .04), but not in either hippocampus (*P* = .14) or PHC (*P* = .65). Error bars indicate ± 1 SEM. **P* < .05.

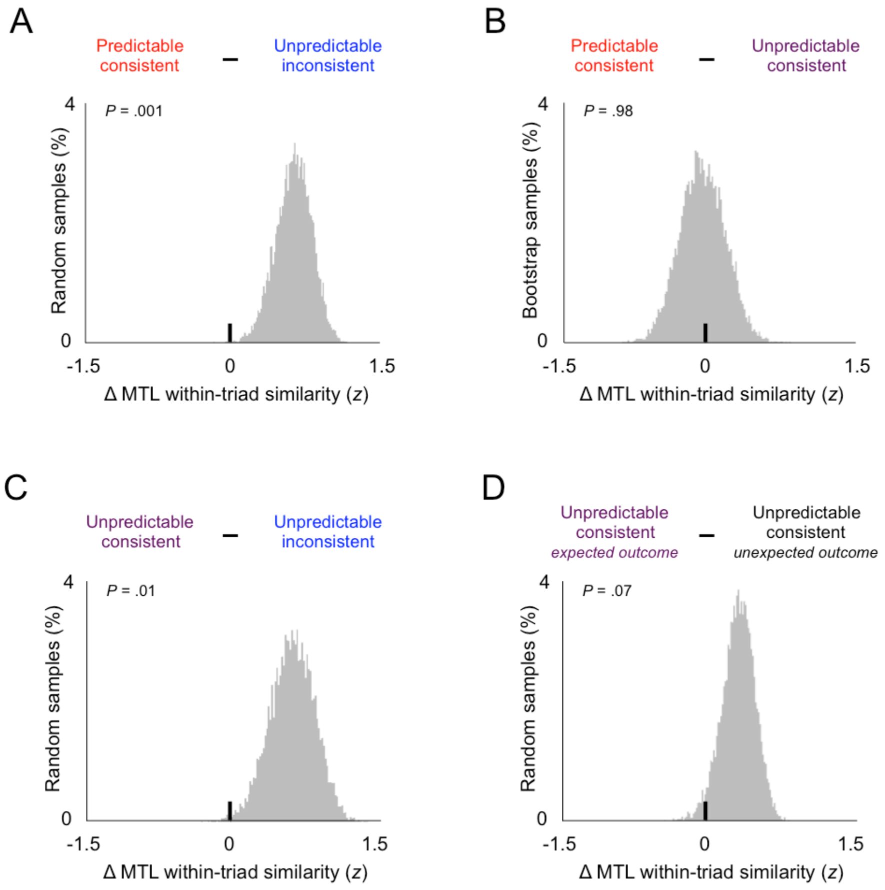

Figure S7. Resampled test consistency. Subject-level bootstrap resampling (Efron and Tibshirani 1986) was used to test the random-effects significance of the relationship between test consistency and MTL (defined as PRC/ERC) within-triad similarity. We resampled with replacement among all 24 subjects for 10,000 random samples. For each random sample, we calculated the difference of within-triad similarity between each type of triad. Across samples, **(A)** MTL within-triad similarity was reliably greater for predictable-consistent triads than for unpredictable-inconsistent triads, but **(B)** did not reliably differ between unpredictable-consistent and unpredictable-inconsistent triads. **(C)** Within-triad similarity was also reliably greater for unpredictable-consistent triads than for unpredictable-inconsistent triads. **(D)** Although primary analyses of unpredictable-consistent triads focused on trials that contained the outcome that the participant subjectively expected to appear, we additionally compared such trials with expected outcomes to trials with the same cues and actions but with unexpected outcomes. Within-triad similarity was marginally greater for unpredictable trials with expected outcomes than for unpredictable trials with unexpected outcomes.

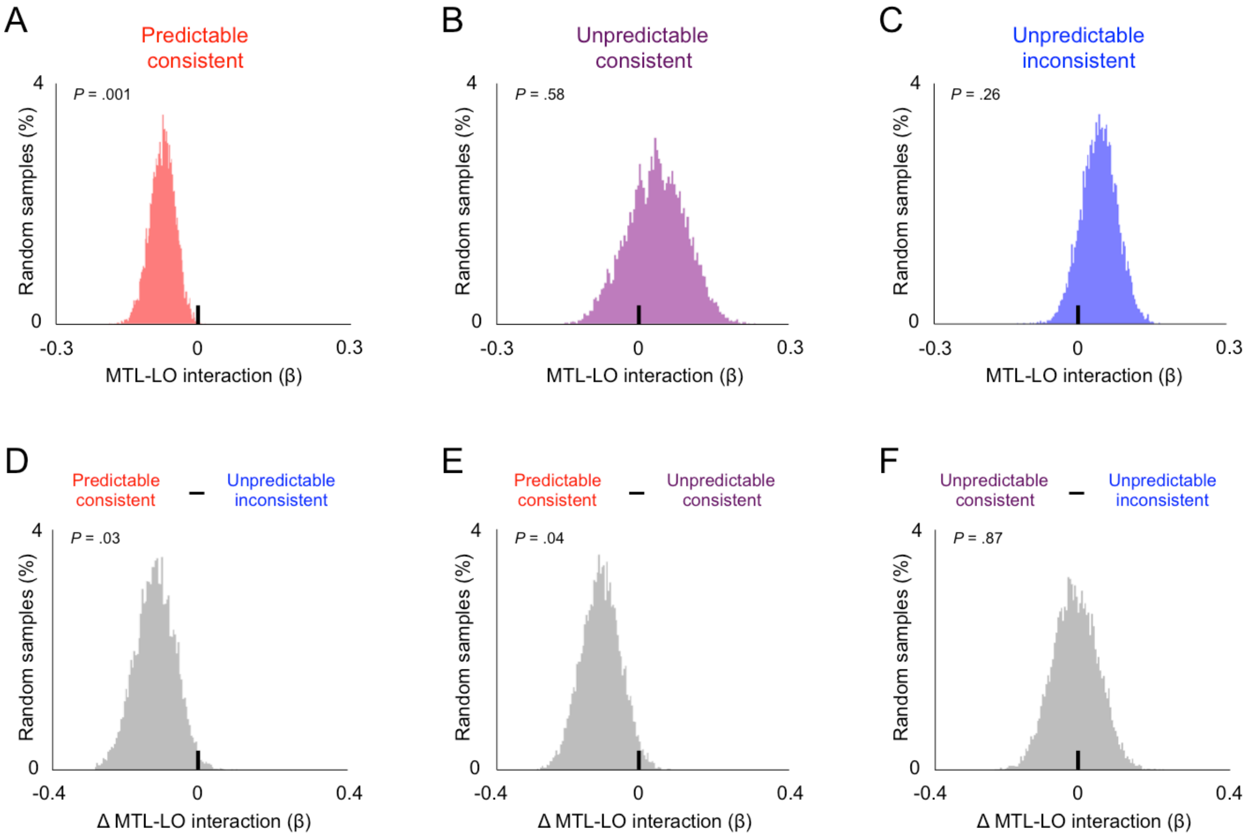

Figure S8. Resampled MTL-LO interaction. Subject-level bootstrap resampling (Efron and Tibshirani 1986) was used to test the random-effects significance of the trial-by-trial relationship between triad reactivation in MTL (defined as PRC/ERC) and overall activity in LO. There was a reliably negative relationship between MTL triad reactivation and LO activity for **(A)** predictable-consistent triads, but no relationship for either **(B)** unpredictable-consistent or **(C)** unpredictable-inconsistent triads. By calculating differences between conditions for each bootstrap sample, we further found that the relationship between MTL triad reactivation and LO activity was reliably different between **(D)** predictable-consistent and unpredictable-inconsistent triads, and between **(E)** predictable-consistent and unpredictable-consistent triads, but did not differ between **(F)** unpredictable-consistent and unpredictable-inconsistent triads.

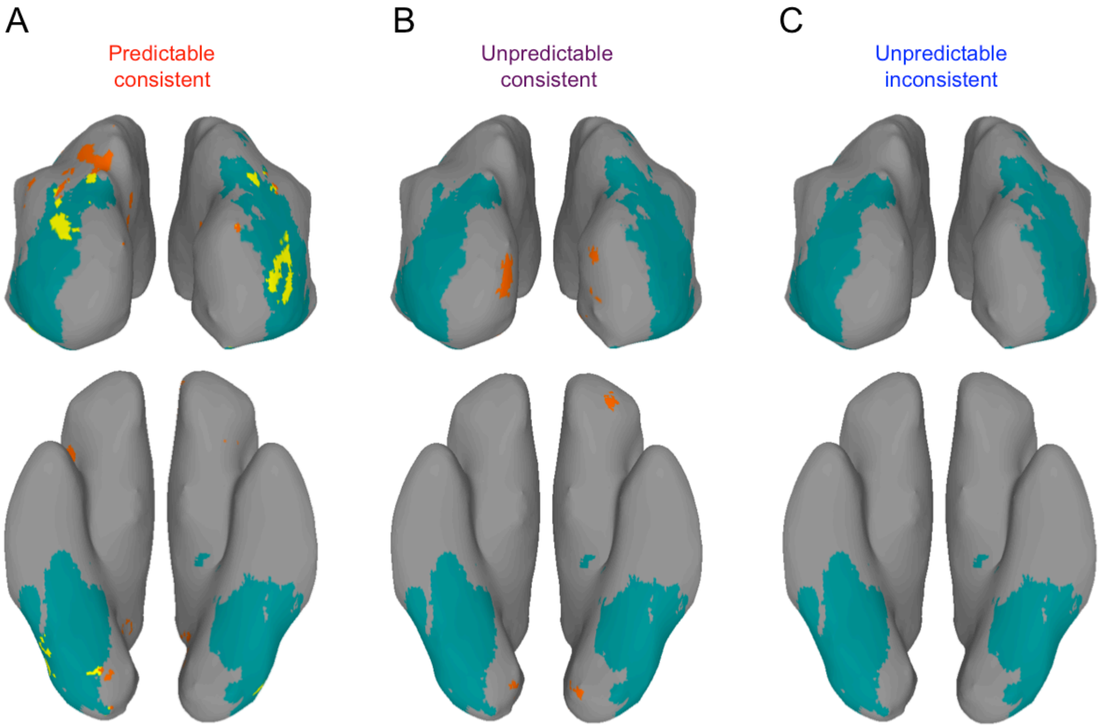

Figure S9. Whole-brain relationship between MTL-cortical interactions and stimulus selectivity. Subject-level bootstrap resampling of the entire brain was used to test the random-effects significance of the negative relationship between PRC/ERC similarity and voxel activity. Specifically, we calculated the trial-by-trial relationship between triad reactivation in MTL and the level of activity of every voxel in the brain for 10,000 bootstrap samples. We then measured the overlap (yellow) between voxels for which MTL triad reactivation negatively predicted activity (orange) and voxels that were deemed stimulus-selective by the localizer (green). **(A)** For predictable-consistent triads, 22.9% of voxels with reliable MTL-cortical interactions were also reliably stimulus-selective. **(B)** Among voxels with reliable MTL-cortical interaction for unpredictable-consistent triads, only 3.2% of were stimulus-selective, a significantly lower proportion than observed for predictable-consistent triads ($\chi^2(1, N = 24) = 15.35$, $P < .001$). **(C)** Triad reactivation did not reliably predict attenuation anywhere in the brain for unpredictable-inconsistent triads. The activity map for each contrast was corrected for multiple comparisons across the whole-brain at $P < .05$, based on a voxelwise α of $P < .001$ and a cluster-forming threshold of 25 voxels determined using 3dClustSim (Cox 1996); http://afni.nimh.nih.gov/pub/dist/doc/program_help/3dClustSim.html).

Supplemental References

Cox RW. 1996. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res. 29:162–173.

Efron B, Tibshirani R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci. 1:54–75.