Prediction strength modulates responses in human area CA1 to sequence violations

Janice Chen,1 Paul A. Cook,1 and Anthony D. Wagner1,2
1Department of Psychology, Stanford University, Stanford, California; and 2Neurosciences Program, Stanford University, Stanford, California

Submitted 16 February 2015; accepted in final form 8 June 2015

Chen J, Cook PA, Wagner AD. Prediction strength modulates responses in human area CA1 to sequence violations. J Neurophysiol 114: 1227–1238, 2015. First published June 10, 2015; doi:10.1152/jn.00149.2015.—Emerging human, animal, and computational evidence suggest that, within the hippocampus, stored memories are compared with current sensory input to compute novelty, i.e., detecting when inputs deviate from expectations. Hippocampal subfield CA1 is thought to detect mismatches between past and present, and detected novelty is thought to modulate encoding processes, providing a mechanism for gating the entry of information into memory. Using high-resolution functional MRI, we examined human hippocampal subfield and medial temporal lobe cortical activation during prediction violations within a sequence of events unfolding over time. Subjects encountered sequences of four visual stimuli that were then reencountered in the same temporal order (Repeat) or a rearranged order (Violation). Prediction strength was manipulated by varying whether the sequence was initially presented once (Weak) or thrice (Strong) prior to the critical Repeat or Violation sequence. Analyses of blood oxygen level-dependent signals revealed that task-responsive voxels in anatomically defined CA1, CA23/dentate gyrus, and perirhinal cortex were more active when expectations were violated than when confirmed. Additionally, stronger prediction violations elicited greater activity than weaker violations in CA1, and CA1 contained the greatest proportion of voxels displaying this prediction violation pattern relative to other medial temporal lobe regions. Finally, a memory test with a separate group of subjects showed that subsequent recognition memory was superior for items that had appeared in prediction violation trials than in prediction confirmation trials. These findings indicate that CA1 responds to temporal order prediction violations, and that this response is modulated by prediction strength.

associative novelty; comparator; fMRI; declarative memory; mismatch detection

Prediction is at the core of memory. By definition, a memory system stores information about the past in service of preparation for the future. As an organism explores the world and encounters stimuli, its state is modified in a way that prepares it for the next interaction with those stimuli. An inherent property of memory systems is the expectation that events will repeat; thus each encounter generates expectations, based on past experience, that the next encounter will resemble prior events. When expectations are confirmed, future predictions will take the same form. When expectations are violated, future predictions are modified accordingly to improve accuracy. The act of learning by error correction necessarily includes both the generation of predictions and the detection of mismatches between expected events and actual experiences/outcomes, so that the organism can learn through experience and update its state appropriately.

Converging human, animal, and computational evidence suggest that, within the hippocampus, stored memories are compared with current sensory input to compute novelty, i.e., detecting when expectation deviates from outcome (Hasselman and Wyble 1997; Kumaran and Maguire 2006, 2007b, 2009; Lisman and Grace 2005; Vinogradova 2001). The emerging theories are strongly guided by the anatomy of the hippocampus (Insauti and Amaral 2004), wherein the dominant information flow begins with entorhinal cortical inputs to dentate gyrus (DG), which projects to CA3. CA3, in turn, projects to CA1, with CA1 inputs to subiculum serving as the dominant output pathway of the hippocampus. Entorhinal cortex (ERC) also provides direct cortical inputs to CA1 (along with to CA3), with this pathway providing a unique source of information to CA1 beyond that arriving from CA3. Given this architecture, it has been hypothesized that CA1 supports computations of mismatch between past and present, i.e., prediction violations; detection of prediction violations in turn is hypothesized to modulate encoding processes, providing a mechanism for gating the entry of information into long-term memory (Hasselman and Wyble 1997; Lisman and Grace 2005).

To test this hypothesis in humans, initial studies used functional MRI (fMRI) and explicit or implicit memory paradigms. In explicit memory paradigms, subjects are trained on item-item associations and later asked to retrieve the associations and make judgments about expected outcomes (associative matches) vs. unexpected outcomes (associative mismatches). Such studies yielded mixed results, sometimes finding hippocampal activity enhancement for associative mismatches compared with associative matches (Kohler et al. 2005), sometimes match enhancement (Hannula and Ranganath 2008), and sometimes both (Duncan et al. 2009). A possible explanation for the mixed results is that these studies used standard resolution fMRI and were thus unable to differentiate the CA1 subfield from the rest of the hippocampus. However, studies using high-resolution fMRI with explicit memory paradigms also report a mixture of mismatch enhancement (Chen et al. 2011; Duncan et al. 2012; Lacy et al. 2011) and match enhancement (Dudukovic et al. 2011; Duncan et al. 2012) in CA1.

By contrast, other studies utilized paradigms in which subjects do not make explicit match/mismatch judgments and instead perform sequential learning and implicit retrieval, often during an unrelated task (e.g., a one-back judgment), along with fMRI (Kumaran and Maguire 2006, 2007b; Lacy et al.
2011) or electrocorticography (Axmacher et al. 2010; Chen et al. 2013). These studies have observed hippocampal enhancement when the present deviates from the past [either greater blood oxygen level-dependent (BOLD) activity or low-frequency power], suggesting that explicit judgments may somehow influence the subjects’ strategies or introduce noise to the measurement of hippocampal prediction violation signals. Indeed, while Duncan et al. (2012) utilized an explicit memory retrieval and judgment task, CA1 activity was enhanced for stimulus changes that were irrelevant to the match/mismatch judgment. Chen et al. (2013) further demonstrated that changes in hippocampal slow-theta power are specific to the onset of prediction-violating stimuli and emerge within 500 ms of such violations. These findings are compatible with the hypothesis that the hippocampus detects violations of associative predictions and suggest that hippocampal mismatch (prediction violation) detection can occur under incidental or implicit conditions. However, no prior study has specifically linked associative prediction violation effects to CA1 in the absence of explicit memory judgments.

Given that hippocampal CA1 responses are proposed to index violations of predictions, a natural question is whether the magnitude of the response is related to the magnitude of the violation. That is, do violations of strong predictions result in greater responses than violations of weak predictions, i.e., a scalable signal? Or is the prediction violation response insensitive to the strength of the prediction? To address this central question, we used a sequential learning task in which prediction strength (memory strength) was manipulated prior to the critical prediction violation event by varying the number of previous exposures. Subjects encountered sequences of four visual stimuli that were then reencountered in the same sequential order (Repeat) or a temporally rearranged order (Violation). Prediction strength was manipulated by varying whether the sequence was initially presented once (Weak) or thrice (Strong) prior to the critical Repeat or Violation sequence. The paradigm did not include any explicit memory judgments; instead, subjects performed an unrelated one-back detection task (Kumaran and Maguire 2006). We used high-resolution fMRI to examine effects in anatomically defined subfields CA1, CA3/DG, and subiculum, along with perirhinal cortex (PRc), parahippocampal cortex (PHc), and ERc. We hypothesized that violations of stronger (relative to weaker) predictions would elicit a larger response in the hippocampus, specifically in CA1. Moreover, we conducted a postscan memory test, examining whether subsequent recognition memory was superior for items that had appeared in trials where predictions were violated than in trials where predictions were confirmed.

MATERIALS AND METHODS

Subjects (fMRI)

Twenty-eight right-handed, native-English speakers (9 women; 18–31 yr old) were recruited from the Stanford University community. Subjects were paid $20/h for their participation. Informed consent was obtained in a manner approved by the Stanford University institutional review board. Data from five subjects were excluded from analysis due to equipment failures, leading to loss of more than 10% of the functional data; an additional two subjects did not complete the study due to discomfort; and data from one subject were excluded because of image quality problems. As such, analyses were performed over data from 20 subjects.

Behavioral Procedures (fMRI)

The behavioral paradigm was modeled after that of Kumaran and Maguire (2006), modified to include a prediction strength manipulation. Subjects performed a 1-back task while viewing object stimuli (color photographs of a variety of objects on white background) in the scanner (Fig. 1). Each trial consisted of four images (a “quartet”) presented one at a time, each appearing for 900 ms with a 100-ms fixation cross interstimulus interval, and ending with a 1,700-ms fixation cross for a total of 5.6 s. The first presentation of a given quartet of images is referred to as a “First” quartet. One-half of all First quartets were followed immediately by a second quartet of the same images in the same order A-B-C-D (“Weak Repeat”), while the other one-half were followed immediately by a quartet of the same images in the rearranged temporal order A-B-D-C (“Weak Violation”). Quartets from Weak Repeat trials were presented again, ~10 min later, in the same order A-B-C-D (“Third”). One-half of all Third quartets were followed immediately by a quartet of the same images in the same order A-B-C-D (“Strong Repeat”), and the other one-half by a quartet of the same images in the rearranged order A-B-D-C (“Strong Violation”). Thus, for “Weak” conditions, quartets were presented one time before being repeated (Weak Repeat) or rearranged (Weak Violation), while in the “Strong” conditions, quartets were presented three times before being repeated for a fourth time (Strong Repeat) or rearranged (Strong Violation). All Weak Repeat images were reiterated in either Strong Repeat or Strong Violation trials, but never in both; except for this, all images were trial-unique.

The remaining trials (22%) contained targets for the 1-back task (a picture identical to the immediately preceding picture), with the target appearing with equal probability in the second, third, or fourth position in the quartet. One-half of these 1-back quartets were immediately preceded by a First quartet of the same pictures, so that targets appeared equally often in First and second quartets. Subjects were instructed to press one of two keys to indicate whether it was or was not identical to the immediately preceding picture; they were told that pictures would appear in sets of four and that targets would always occur within (never across) sets. A 0- to 19.6-s intertrial interval followed each trial, during which subjects viewed a fixation cross, and the order of conditions...
optimized efficiency of the event-related fMRI design (Dale 1999). Each experimental run (~7 min) contained 54 quartets: 18 First, 6 Weak Repeat, 6 Weak Violation, 6 Third, 3 Strong Repeat, 3 Strong Violation, and 12 1-back (6 of which were preceded by a First quartet). Items used in 1-back trials were trial-unique (within subjects, they did not appear in other conditions or in other 1-back trials). Images were assigned randomly to conditions, with a new random assignment generated for each subject.

At the end of each run, subjects were shown their percent accuracy on the 1-back task for that run. All subjects completed 12 runs of the experiment, except for one subject who completed 11 runs. Prior to the 12 experimental runs, an initial run of trials was presented (also scanned) which did not contain any Strong trials, consisting only of six Weak Repeat, six Weak Violation, and six Target trials. This initial run was needed to generate Strong trials for the first experimental run; neither the behavioral nor imaging data from the initial run were included in the analyses.

A memory test was administered outside of the scanner ~15 min after the end of the sequence task. The test was composed of all pictures that had occurred in the third or fourth positions in Weak Violation, Strong Repeat, and Strong Violation trials (144 trials, 288 pictures total), and an equal number of novel foils; stimuli were presented in random order. (Weak Repeat trials could not be tested because all Weak Repeat pictures were reiterated in Strong Repeat and Strong Violation trials.) On each trial, subjects viewed one picture and reported their confidence that it was old (recognized) or new on a four-point scale: 1, “Definitely Old”; 2, “Probably Old”; 3, “Probably New”; 4, “Definitely New”.

Stimuli were presented and responses recorded using MATLAB (Mathworks) and the Psychophysics Toolbox (Brainard 1997; Pelli 1997). Responses during the sequence task were made with the right hand using a scanner-compatible button box, and the memory test was administered on a laptop computer.

fMRI Procedures

fMRI data were acquired on a 3.0T GE Signa whole body system with a quadrature transmit/receive head coil. Head movement was minimized using foam padding. Before functional imaging, high-resolution T2-weighted spin-echo structural images [repetition time (TR) = 3,000 ms; echo time (TE) = 68 ms; 0.43 × 0.43 mm in-plane resolution] were acquired in 20 2-mm-thick slices parallel to the main axis of the hippocampus (axial-obliques). After functional imaging, a second set of T2-weighted images was collected with the same parameters but oriented perpendicular to the hippocampal axis (coronal-obliques), allowing for the segmentation of hippocampal subfields and medial temporal lobe (MTL) cortical areas in the coronal plane; these regions of interest (ROIs) were then coregistered and resliced to the axial plane for application to the functional data.

A total of 1,812 (12 scans, 19 subjects) or 1,661 (11 scans, 1 subject) functional volumes were acquired for each subject using a T2*-sensitive gradient echo spiral in/out pulse sequence (Glover and Law 2001) with the same slice locations as the axial-oblique structural images (TR = 2,800 ms; TE = 30 ms; flip angle = 61°; two shots; field of view = 22 cm; 2.0 × 2.0 × 2.0 mm resolution). A high-order shimming procedure, based on spiral acquisitions, was used to reduce B0 heterogeneity (Kim et al. 2002). The first three volumes of each scan (a total of 8.4 s) were discarded to allow for T1 stabilization.

Imaging Analyses

Data were slice-time and motion corrected (realigned to the first volume in the time series) using SPM5 (http://www.fil.ion.ucl.ac.uk/spm/) and custom MATLAB routines. A mean T2*-weighted volume was computed during realignment, and the T2*-weighted axial and coronal anatomical volumes were coregistered to this mean functional volume. Data were not smoothed or normalized. A statistical model was constructed for each subject in which separate event regressors were defined for First, Weak Repeat, Weak Violation, Third, Strong Repeat, Strong Violation, and 1-back (containing a target) quartets. The model also included regressors of no-interest that accounted for effects of scan session and linear drift.

For each subject, anatomical ROIs were manually demarcated on the high-resolution coronal structural images (Fig. 2) using established procedures (Insauti et al. 1998; Insauti and Amaral 2004; Pruessner et al. 2000, 2002; Zeineh et al. 2000, 2003; for recent discussion of hippocampal segmentation, see Yushkevich et al. 2015). We demarcated hippocampal subfields [CA1, CA2/3/DG, and subiculum; note that DG, CA2, and CA3 cannot be unambiguously segmented at the acquired resolution (Carr et al. 2010)], the remaining anterior and posterior portions of hippocampus that could not be definitively decomposed into separable hippocampal subfields, and MTL cortical areas (Prc, Phc, and ERe). To further constrain analyses to voxels that were at least modestly responsive to task demands, functional-anatomical ROIs were created by defining voxel-level contrast images at a liberal threshold (t > 0.675; P < 0.25, no cluster minimum size) masked with the anatomical ROIs for each subject. These functional-anatomical ROIs were always created using a contrast independent of the conditions of interest (specifically, from the same statistical model as described above, a 1-back > baseline contrast was used to create functional ROIs, within which responses to Repeat and Violation trials were then examined). For each anatomical ROI and each functional-anatomical ROI, percent signal change was extracted for each subject with MarsBaR (Brett et al. 2002), using a finite impulse response model to estimate the response at each TR and then summing over TRs 3–5 (i.e., 5.6–14.4 s after quartet onset). The TRs 3–5 window choice reflects the fact that the critical event (sequence rearrangement at the third item in Violation quartets) occurs at 2.0 s postquartet onset. That is, as the A and B items are identical to the A and B items from the immediately prior sequence, the Repeat and Violation conditions do not differ until the third position in the sequence. While the signal at TRs 3–5 may reflect effects of earlier items, the differences between conditions in the signal can only be due to the third and fourth position items. The same TRs 3–5 window was used for analyses of repetition effects. Sum percent signal change was submitted to repeated-measures ANOVA, treating subjects as a random effect; hemisphere was included as a factor in all ANOVAs.

Subjects (Behavioral Experiment)

Thirty-five native-English speakers were recruited from the Stanford University community and paid $10/h for their participation.

Procedures (Behavioral Experiment)

A behavioral-only experiment was conducted with a separate group of subjects. We reasoned that the change in environment from scanner to postscan test (outside the scanner) and the overall length of the test may have decreased our sensitivity to differences in memory between conditions. The design of the behavioral experiment consisted of a sequence presentation (encoding) session (Fig. 1) and a memory test that were identical to the fMRI experiment in terms of timing and stimuli, with the following key changes.

1) The item recognition memory test consisted of “C” and “D” items from Strong Repeat and Strong Violation quartets only, with an equal number of foils. (In the above postscan memory test, items from Weak quartets were included as well.) We reasoned that the length of the test when including Weak quartet items, and interference from the additional trials, might have increased over-
all difficulty and reduced our sensitivity to differences in memory for Strong quartet items. Thus the Weak condition was still present in the encoding portion of the behavioral experiment, but these items did not appear in the memory test.

2) An associative recognition memory test (two-alternative forced choice design) was added, in which subjects were cued with “B” items and asked to select items which had originally appeared in the same quartet with the cue. Each trial consisted of three simultaneously presented items: the “B” (cue) item at the top of the screen, and the target and foil below to the left and right (target side was randomized). Targets consisted of “C” and “D” items, in different trials (so that each “B” item appeared twice in the associative memory test). Subjects responded as to whether the target was either 1, “Definitely Left”; 2, “Probably Left”; 3, “Probably Right”; or 4, “Definitely Right”. To generate suitable foils (matched in the number of previous exposures to the targets), the encoding session included some trials on which only a single item appeared (“Solo”). Each Solo item appeared four times over the course of the encoding session (same as Strong Repeat and Strong Violation items). The associative memory test was always administered last, after the item recognition memory test.

RESULTS

FMRI Behavioral Performance

During scanning, subjects successfully responded to 1-back items 87.6% (SD 10.2%) of the time on average, with reaction times of 684 ms (SD 195 ms) on average. The postscanning memory test consisted of all “C” and “D” items from Weak Violation, Strong Repeat, and Strong Violation quartets, with an equal number of foils (Table 1). Accuracy was significantly above chance [mean \(d' = 1.22, SD = 0.46; t_{(19)} = 2.4, P < 0.05\)], with memory significantly better for Strong Repeat \([F_{(1,19)} = 158.83, P < 5 \times 10^{-10}]\) and Strong Violation \([F_{(1,19)} = 95.23, P < 5 \times 10^{-8}]\) items than for Weak Violation items. Similarly, reaction times in the memory test were significantly faster for Strong Repeat \([F_{(1,19)} = 16.34, P < 0.005] \) items than for Weak Violation items. There were no significant differences in accuracy for Strong Repeat vs. Strong Violation items \([F_{(1,19)} = 2.18, P > 0.1] \), nor for...
values are means (SD). RT, reaction time.

“C” vs. “D” items (F < 1), and there was no interaction (F < 1). There were no significant differences in reaction time for Strong Repeat vs. Strong Violation items (F < 1), nor for “C” vs. “D” items [F(1,19) = 1.00, P > 0.1], and no interaction (F < 1).

Memory for Violation Items Is Enhanced in a Separate Group of Subjects

To further explore consequences for memory in the sequence task, we conducted a separate behavioral-only experiment in which memory was tested only for Strong quartet items. We reasoned that, when Weak quartet items were included in the postscan test, the length of the test and interference from the additional trials might have increased overall difficulty and reduced our sensitivity to differences in memory for Strong quartet items. Thus the Weak condition was still present in the encoding portion of the separate behavioral-only experiment, but these items did not appear in the memory test (although, given enough power, one might predict memory enhancement for Weak Violation items).

In this simplified item recognition memory test (Table 2), a significant main effect of Condition [F(1,34) = 5.6, P < 0.05] was observed, revealing that the hit rate was modestly, but significantly, higher for Violation items than for Repeat items (no main effect of C/D or interactions, F values < 1). Incorrect RTs were longer than correct RTs [F(1,32) = 10.5, P < 0.005; note: two subjects were removed from RT analyses due to having no Incorrect trials in some conditions]. In the associative recognition memory test, neither accuracy nor RT significantly differed between conditions [F values < 1; Table 3].

Repetition and Prediction Violation Effects in Functional-Anatomical MTL ROIs

Functional-anatomical ROIs were created by restricting anatomical ROIs to voxels that were at least modestly engaged during the task compared with baseline. As the conditions of interest were Repeat/Violation and Weak/Strong object quartets, voxels were selected using an independent set of trials: all quartets containing 1-back targets. The contrast of these 1-back trials > baseline was thresholded at t > 0.675 (P < 0.25 uncorrected, no cluster minimum size) and masked with the anatomical ROIs to create the functional-anatomical ROIs. The resulting functional-anatomical ROIs (hereafter referred to as “functional ROIs” for simplicity) consisted of 20–55% of the voxels in the anatomical ROIs.

Given our strong a priori hypotheses about the role of CA1 in prediction violation detection, all statistical tests are reported here without correction for multiple comparisons. However, for hippocampal subregions, we indicate which tests remain significant (**P < 0.05) or show a trend (*P < 0.1) when using Bonferroni correction over the three hippocampal subfield ROIs tested (for CA1, CA23/DG, subiculum). As we did not have strong a priori hypotheses regarding MTL cortex, for these regions we indicate which tests survive correction over the eight hippocampal and MTL cortical ROIs tested [the three hippocampal subfields, plus anterior hippocampus, posterior hippocampus, PRc, PHc, and Erc (**P < 0.05; *P < 0.1)].

Repetition effects. The effect of simple repetition was examined by comparing trials with non-rearranged quartets: First, Weak Repeat, Third, and Strong Repeat (Fig. 1). As Weak Repeat trials consist of quartets being shown for the second time, and Strong Repeat trials consist of quartets being shown for the fourth time, the four trial types can be described as presentation 1, presentation 2, presentation 3, and presentation 4 (Fig. 3). Importantly, presentations 1 and 2 were consecutive in time, and presentations 3 and 4 were consecutive, but presentations 2 and 3 were separated by ~10 min (see MATERIALS AND METHODS). Integrated percent signal change for each functional ROI was submitted to ANOVA with immediate repetition (presentations 1 and 3/presentations 2 and 4), lagged repetition (presentations 1 and 2/presentations 3 and 4), and hemisphere included as factors. The effect of immediate repetition was significant in PRc [F(1,19) = 37.36, **P < 0.0005] and PHc [F(1,19) = 77.71, ***P < 0.0005], with activity lower for presentation 2 than presentation 1, and lower for presentation 4 than presentation 3, i.e., repetition suppression [Fig. 3; all other ROIs, F(1,19) < 3.2, P > 0.09]. The effect of lagged repetition was significant in CA1, CA23/DG, PRc, PHc,

Table 2. Performance on the behavioral-only item recognition memory test

<table>
<thead>
<tr>
<th></th>
<th>“C” Item</th>
<th>“D” Item</th>
<th>False Alarm Rate</th>
<th>Correct Rejection RT, s</th>
<th>False Alarm RT, s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hit rate</td>
<td>Hit RT, s</td>
<td>Miss RT, s</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.76 (0.16)</td>
<td>1.12 (0.24)</td>
<td>1.31 (0.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.79 (0.15)</td>
<td>1.08 (0.24)</td>
<td>1.37 (0.76)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.76 (0.14)</td>
<td>1.09 (0.25)</td>
<td>1.33 (0.47)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.76 (0.16)</td>
<td>1.08 (0.22)</td>
<td>1.38 (0.61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.79 (0.15)</td>
<td>1.07 (0.23)</td>
<td>1.22 (0.71)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.16 (0.12)</td>
<td>1.48 (0.50)</td>
<td>1.24 (0.33)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Values are means (SD).
posterior hippocampus \( F_{(1,19)} > 11, **P < 0.005 \), subiculum \( F_{(1,19)} = 8.03, **P = 0.01 \), and a trend in ERc \( F_{(1,19)} = 4.12, P = 0.06 \), but not in anterior hippocampus \( F < 1 \).

**Violation effects.** The effect of sequence rearrangement was examined by comparing trials with rearranged and non-rearranged quartets. Integrated percent signal change for each functional ROI was submitted to ANOVA with Repeat/Violation, prediction strength (Weak/Strong), and hemisphere included as factors. As hypothesized, the effect of Repeat/Violation was significant in CA1 \( F_{(1,19)} = 8.77, **P < 0.01 \), with greater activity on Violation than Repeat trials (Fig. 4). There also was a significant effect in PRc \( F_{(1,19)} = 5.05, P < 0.05 \), and trends in CA23/DG, ERc, and PHc \( P \) values \( < 0.1 \). The effect of prediction strength was significant in CA1 \( F_{(1,19)} = 5.70, *P < 0.05 \), CA23/DG \( F_{(1,19)} = 11.04, **P < 0.005 \), PHc \( F_{(1,19)} = 34.38, **P < 0.00005 \), PRc \( F_{(1,19)} = 5.51, P < 0.05 \), subiculum \( F_{(1,19)} = 5.82, P < 0.05 \), and posterior hippocampus \( F_{(1,19)} = 16.19, **P < 0.001 \), and a trend was found in ERc \( P < 0.1 \), with greater activity on Weak trials than Strong trials. Critically, as hypothesized, an interaction between prediction strength and Repeat/Violation was found in CA1 \( F_{(1,19)} = 4.58, P < 0.05 \); all other regions \( P \) values \( > 0.1 \), stemming from significantly greater activity for Violation than Repeat in the Strong condition \( F_{(1,19)} = 7.33, **P = 0.01 \) but not the Weak condition \( F < 1 \). Although PRc displayed a qualitatively similar pattern \( [Strong Violation > Strong Repeat, F_{(1,19)} = 4.28, P = 0.05; Weak Violation vs. Repeat, F < 1] \), the interaction between prediction strength and Repeat/Violation was not significant \( P > 0.2 \). However, the region by Repeat/Violation interactions between CA1 and CA23/DG, PRc, or PHc were not significant \( F \) values \( < 1 \).

Table 3. **Performance on the behavioral-only associative recognition memory test**

<table>
<thead>
<tr>
<th></th>
<th>“C” Item</th>
<th></th>
<th>“D” Item</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hit rate</td>
<td>Hit RT, s</td>
<td>Miss RT, s</td>
</tr>
<tr>
<td>Repeat</td>
<td>0.56 (0.12)</td>
<td>1.29 (0.41)</td>
<td>1.24 (0.48)</td>
</tr>
<tr>
<td>Violation</td>
<td>0.57 (0.12)</td>
<td>1.26 (0.41)</td>
<td>1.28 (0.49)</td>
</tr>
</tbody>
</table>

Values are means (SD).

[Note that all results were highly similar when data from anatomical (unrestricted) ROIs were analyzed (see ENDNOTE).]

**Subsequent memory analyses.** We tested for a relationship between the magnitude of CA1 Strong Violation \( > \) Strong Repeat and later recognition accuracy for Violation C and D items in the postscan memory test using across-subject correlation; the relationship was not significant \( (R = -0.05, P = 0.82) \). We also explored the relationship between CA1 repetition suppression and subsequent memory in the postscan test using a new statistical model in which subsequently remembered and forgotten items were divided into separate regressors for Weak Repeat, Strong Repeat, Weak Violation, and Strong Violation trials. This analysis revealed no significant difference between activity for remembered vs. forgotten trials in any of the four conditions \( (P \) values \( > 0.65) \). Next, we examined the correlation between CA1 activity during the remembered conditions with subsequent accuracy for items from Strong Repeat, Weak Violation, and Strong Violation trials, across subjects. There were no significant correlations \( (P \) values \( > 0.13 \). These null findings are perhaps unsurprising given that, in the postscan memory test, we did not find a significant difference between Strong Repeat and Strong Violation items.

**Voxel activation extent.** The contrast of 1-back > baseline was chosen for creation of the functional ROIs to identify voxels that were engaged during the task compared with baseline. To investigate whether the above results depended on the choice of contrast, we created a separate set of functional ROIs using a contrast of all First and Third quartets over baseline (First + Third > baseline). Importantly, the main results did not change: significant main effects of Repeat/Violation were found in CA1 \( F_{(1,19)} = 6.78, P < 0.05 \) and

Fig. 3. Repetition effects in functional ROIs. A–D: integrated percent signal change (TRs 3–5) for repeated (intact order) quartets in CA1 (A), CA23/DG (B), perirhinal cortex (PRc; C), and parahippocampal cortex (PHc; D). E–H: corresponding percent signal change time courses for CA1 (E), CA23/DG (F), PRc (G), and PHc (H). Values are means \( \pm \) SE of subjects. Functional ROIs were created from the contrast of 1-back trials > baseline and thresholded at \( t > 0.675 \).
PREDICATION STRENGTH MODULATES VIOLATION DETECTION IN CA1

More Voxels Were Sensitive to Violation in CA1 than in CA23/DG

Sensitivity to violation was investigated in each region by comparing the distributions of t values in the contrast of Strong Violation > Strong Repeat. For each functional ROI, the number of voxels exceeding threshold was calculated for a series of values ranging from t = 0 to t = 1.282 (i.e., P > 0.5 to P < 0.1) and divided by the total number of voxels for that ROI (Fig. 5A). CA1 consistently contained the greatest proportion of voxels above threshold, regardless of which threshold was used. A t-test showed that the proportion of voxels passing the threshold of t = 0.675 in functionally defined CA1 was significantly greater than in CA23/DG (t(19) = 3.04, P < 0.01]. (As the proportion values at different thresholds are non-independent, we report only the t = 0.675 threshold, the same threshold used for voxel selection for functional ROIs from the independent contrast of 1-back > baseline.) The same calculation was performed using the full anatomical ROIs; again, CA1 consistently contained the greatest proportion of voxels above threshold, regardless of threshold (Fig. 5B). A t-test showed that the proportion of voxels passing the threshold of t = 0.675 was significantly greater in CA1 than in CA23/DG (t(19) = 2.63, P = 0.016]. To determine whether this response advantage in CA1 was a general property or selective to Strong Violation > Repeat, we calculated the number of voxels exceeding threshold (same range as above) in the contrast of 1-back > baseline (the same contrast used to create the functional ROIs) and divided by the total number of voxels for each anatomical ROI (Fig. 5C). The proportion of CA1 voxels was not on average higher than other regions, and in fact was the lowest of all regions for thresholds t = 0.675 and above. That is, the finding that CA1 contained the greatest proportion of voxels responsive to Strong Violation > Strong Repeat was not due simply to general responsiveness.

Finally, to further probe CA1 response properties to Violation, the proportion of voxels above the t = 0.675 threshold was calculated in the contrast of Weak Violation > Weak Repeat for CA1 and CA23/DG functional ROIs and submitted to repeated-measures ANOVA with Weak/Strong and ROI as factors. There was a main effect of ROI indicating that CA1 had a greater proportion of voxels above threshold than CA23/DG for Violation > Repeat, regardless of Weak vs. Strong (F(1,19) = 7.44, P = 0.011. There was no main effect of Weak/Strong (F(1,19) < 1) and no interaction (F(1,19) = 1.2, P > 0.2].

DISCUSSION

BOLD activity was assessed in hippocampal subregions and MTL cortical regions while subjects viewed sequentially presented unique image quartets and performed an incidental 1-back task. The number of exposures to each image quartet was varied to manipulate the subjects’ experience with, and thus the predictability of, a given quartet order. Some quartets were shown one time ("Weak Prediction") and some were...
shown three times (“Strong Prediction”) before appearing in a final repeated (Repeat: ABCD) or rearranged (Violation: ABDC) order.

Within anatomically defined regions, an independent contrast was used to identify voxels that were engaged during the object sequence task. Analyses of BOLD activity in these functional ROIs showed that CA1 and PRc were significantly more responsive during Violation than Repeat trials. Additionally, in CA1 only, there was a significant Prediction Strength × Repeat/Violation interaction, wherein Violation > Repeat was greater on Strong Prediction trials than on Weak Prediction trials. Highly similar results were found using the full anatomical ROIs. Comparison of voxel populations in anatomical and functional ROIs showed that CA1 consistently contained the greatest proportion of voxels displaying greater responses to Violation than Repeat of any ROI. Finally, a separate behavioral experiment showed that recognition memory was better for items that had appeared in Violation than in Repeat trials. Thus the main findings were as follows: 1) enhanced activity for prediction violations (Violation > Repeat) was found in CA1 and PRc; 2) the prediction violation response was modulated by prediction strength in CA1; 3) CA1 contained the highest proportion of voxels displaying enhanced activity for prediction violations of any ROI; and 4) recognition memory was better for items that had appeared in rearranged (Violation) than intact (Repeat) orders.

Prediction Violation Responses in CA1 and PRc

Voxels in CA1 were found to be more responsive to prediction violations (mismatch between prediction and outcome) compared with confirmed predictions. These results converge with previous whole brain studies demonstrating mismatch enhancement in the hippocampus (Kumaran and Maguire 2006, 2007b), and further show that, within the hippocampus, such responses can be observed in the CA1 subfield. Previous studies of CA1 prediction violation responses used paradigms in which subjects made explicit match/mismatch judgments (Chen et al. 2011; Dudukovic et al. 2011; Duncan et al. 2009, 2012; Hannula and Ranganath 2008; Olsen et al. 2009). In the present study, CA1 and PRc prediction violation responses were generated under viewing conditions that did not require subjects to actively retrieve items from memory or evaluate their match/mismatch status. Rather, subjects performed an unrelated task (1-back target detection), and thus the observed responses suggest an automatic process of memory retrieval and prediction violation detection. The involvement of CA1 aligns with anatomically motivated proposals of its function as a novelty detector in the hippocampal circuit, in which predictions generated via pattern completion mechanisms in CA3 are compared with actual outcomes (sensory information arriving from entorhinal cortex) by CA1 (e.g., Lisman and Grace 2005).

The current observation of prediction violation responses in PRc echoes prior work demonstrating a PRc sensitivity to prediction violations when the stimuli are faces (Chen et al. 2011). One possible explanation is that, when an unexpected event is detected in CA1, it triggers an upregulation of encoding processes and a consequent increase in activity in MTL structures responsible for encoding (e.g., Fyhn et al. 2002; Lisman and Grace 2005). PRc might be disproportionately recruited for such encoding due to a bias for objects relative to other types of stimuli; anatomical data from nonhuman animals suggest that PRc differentially receives inputs from neocortical areas that process nonspatial identity of stimuli (i.e., the “what” ventral visual stream), while PHc differentially receives inputs from areas concerned with the spatial content of stimuli (i.e., the “where” dorsal visual stream) (Manns and Eichenbaum 2006). However, some studies have suggested that, in memory paradigms, PRc responds to object and spatial (e.g., scene) stimuli, while PHc preferentially responds to spatial stimuli (Buffalo et al. 2006; Liang et al. 2013; Libby et al. 2014; Preston et al. 2010; Reagh and Yassa 2014).

An alternative interpretation for the prediction violation response in PRc is that it reflects the reinstatement of retrieved representations generated in the hippocampus. In a visual pair-association task, Naya et al. (2001) showed that cue-triggered activity (“perceptual signal”) reached area TE before
an experimental design where the prediction window was generated. A possible way to examine this issue would be with terms of the number of trials on which a prediction was absent vs. present, with Strong and Weak prediction differing predictions) of upcoming items were not Weak vs. Strong, but rather more likely to successfully elicit predictions than were Weak prediction trials. For example, on Repeat trials, B cues the retrieval of C, C cues the retrieval of D, and D cues nothing (having never been immediately followed by a stimulus). On Violation trials, B cues the retrieval of C, D is presented instead, then C is presented in the fourth position; C cues the retrieval of D, and this final retrieval event effectively extends the length of the trial, resulting in a difference in activity between Repeat and Violation trials.

Violation Enhancement in CA1 Is Modulated by Prediction Strength

Studies of reward prediction error learning have shown that the presentation of predictive cues elicits activity in dopamine neurons that is proportional to the magnitude, probability, or delay of the predicted reward (Niv and Schoenbaum 2008). The hippocampal prediction violation response is analogous to a reward prediction error in that it is thought to drive learning based on the mismatch between prediction and outcome. Thus an important question about putative prediction violation (mismatch) responses in CA1 is whether the detection of prediction violation results in a scalable signal. That is, does the strength of the prediction (the strength of the retrieved memory) have any effect on the magnitude of the response?

The design of the present study sought to address this question by manipulating prediction strength via repetition: Weak Prediction trials consisted of quartets for which there had been one previous exposure, while Strong Prediction trials consisted of quartets for which there had been three previous exposures. As hypothesized, Strong Prediction trials generated greater prediction violation responses than Weak Prediction trials in CA1. While this interaction was significant only in CA1, qualitatively similar patterns were observed in PRc and CA23/DG. There were no significant interactions between CA1 and CA23/DG, PRc, or PHc, so the selectivity of this effect should be interpreted with caution. However, supporting the proposal that CA1 plays a primary role in violation detection, CA1 consistently contained the greatest proportion of voxels responsive to Strong Violation > Strong Repeat, regardless of threshold (Figs. 5, A and B). Assessment of functional ROI characteristics in all regions demonstrated that this advantage was not due to general responsiveness (Fig. 5C).

An interesting alternative description of the strength-modulated CA1 violation effect is that Strong Prediction trials were more likely to successfully elicit predictions than were Weak Prediction trials. In other words, perhaps memories (predictions) of upcoming items were not Weak vs. Strong, but rather absent vs. present, with Strong and Weak prediction differing in terms of the number of trials on which a prediction was generated. A possible way to examine this issue would be with an experimental design where the prediction window was separated temporally from the violation window, e.g., as in Chen et al. (2011). If Strong prediction showed more evidence than Weak prediction of pattern reinstatement in the time period prior to the upcoming violation event, on an individual trial basis, this would suggest that prediction strength was truly greater for Strong prediction trials. On the other hand, if Strong and Weak prediction conditions showed similar magnitudes of pattern reinstatement but a greater proportion of such trials was present in the Strong condition, this would suggest that the probability of generating a prediction was greater for Strong than Weak prediction trials.

It is not possible to determine, given the current data, whether the strength-sensitive prediction violation response originates in CA1; nonetheless, the effect appears to be most prominent in CA1, which accords with anatomically motivated theories that CA1’s role within the hippocampal circuit is that of a mismatch detector (Hasselmo and Wyble 1997; Kumaran and Maguire 2007a; Lisman and Grace 2005). The results also align with prior work showing that gradual changes in item appearance (e.g., viewpoint) yield graded changes in CA1 response (Lacy et al. 2011), and that CA1 activity tracks the number of featural changes in a visual scene during match/mismatch judgments (Duncan et al. 2012).

Why Was There No CA1 Prediction Violation Response in the Weak Condition?

The Weak Violation condition in the current study corresponds closely to the condition $S_{mod}$ in Kumaran and Maguire (2006), in which a hippocampal prediction violation response was observed. This raises the question of why the current results do not seem to replicate those of Kumaran and Maguire (2006), with only the Strong Violation condition demonstrating activity enhancement in hippocampal CA1. The answer might simply be that the finding is a null result due to insufficient power (high-resolution fMRI vs standard-resolution fMRI) or slightly different task parameters (e.g., timing: Kumaran and Maguire 2006 used 1,000-ms images with a 200-ms interstimulus interval, while the current study used 900-ms images with a 100-s interstimulus interval). Alternatively, we note that a pattern conceptually similar to the current results is seen in Duncan et al. (2012), wherein the number of associative violations (“changes”) on each trial was parametrically manipulated (rather than the strength of past predictions, as in the current study). Duncan et al. (2012) show a significant linear trend, with more changes eliciting more CA1 activity. However, the two “weakest” violation conditions (one and two changes) would be unlikely to statistically exceed the “no change” condition, with the “one change” condition actually eliciting numerically lower CA1 activation than the “zero changes” condition. An interesting possible explanation that incorporates both sets of results is that the CA1 prediction violation response is context dependent, with the response being primarily elicited by the task condition that is the most unexpected in the context of each experiment. Future studies could explore this hypothesis by comparing blocks of trials with different levels of violation strength intermingled, to test whether the CA1 response to what is called a “strong” violation in the current study is reduced when an even stronger violation condition is present.

J Neurophysiol • doi:10.1152/jn.00149.2015 • www.jn.org
The Role of Temporal Information in Prediction Violation Detection

Recent fMRI studies have revealed that activity in the human hippocampus carries information about the temporal order of items in a sequence (Hsieh et al. 2014). Hippocampal activity also can be used to predict subsequent memory for temporal order (DuBrow and Davachi 2014), and activity changes over the course of learning to reflect temporal relationships between items (Schapiro et al. 2012). In the rodent, time cells in the hippocampus have been described that appear to track the passage of time (MacDonald et al. 2011). In the current design, items were presented individually, allowing predictions to be generated as the sequences unfolded over time. The timing of prior sequence exposures was identical to that of the critical violation trials. Given the richness of temporal information present in hippocampal activity, a potential question is how sensitive/robust hippocampal prediction violation responses are to variations in timing. In real life, repeated event sequences do not always have identical timing (e.g., how much time passes between dialing a phone number and having the call answered), so one might expect that some flexibility in predictive timing would be beneficial. On the other hand, similar sequences may be differentiated purely by their timing (e.g., deciding whether to hold the door for someone depends on estimating how long it will take for them to arrive). An interesting avenue of future research would be to examine how the hippocampus generates predictions and detects prediction violations in situations with variable temporal properties.

Repetition Suppression

Throughout the MTL, responses decreased with multiple repetitions of unique quartets. Only MTL cortical regions (PRc and PHc) showed decreased responses for immediate repetition (two identical consecutive quartets), while both MTL cortex and most hippocampal regions showed decreased responses for lagged repetition (two identical quartets separated by ~10 min).

The current design does not strictly allow a distinction to be made between item- and association-driven repetition suppression. Nonetheless, the pattern of suppression in hippocampal (lagged repetition only) vs. MTL cortical regions (both immediate and lagged repetition) may suggest that MTL cortical regions were sensitive to the immediate repetitions of items, while associations required more exposures to learn and thus hippocampal regions reflected this learning later. Repetition suppression effects have been previously observed in the MTL, with MTL cortical regions tending to be sensitive to item repetitions (Henson 2005; Henson et al. 2003; O’Kane et al. 2005), and hippocampus tending to be sensitive to repetitions of item-item associations (Kohler et al. 2005; Mayes et al. 2007).

The lagged repetition effect helps to elucidate the pattern of results in the Violation condition of CA1 (Fig. 4, A and E) and to a lesser extent in CA23/DG, PRc, and PHc (Fig. 4, B–D and F–H). Whereas the absolute magnitude of Weak Violation and Strong Violation responses do not differ (Fig. 4E), Strong Repeat (fourth presentation of identical quartet) is lower than Weak Repeat (second presentation of identical quartet), and the elevation of Violation above Repeat is greater when predictions are stronger, i.e., Strong Violation minus Strong Repeat is significantly greater than Weak Violation minus Weak Repeat. All else being equal, one should expect that the fourth repetition of items in the Strong Violation condition would be suppressed to the same degree (equally as low) as in the Strong Repeat condition; the elevation of the Strong Violation activation thus stemmed from the temporal reordering of the last two items in the sequence.

The Relationship Between Prediction Violation and Subsequent Memory Performance

Memory was tested for C and D items in a postscan recognition memory test. Accuracy was higher for Strong Prediction quartet items than for Weak Prediction quartet items, as expected, given the greater number of exposures for Strong Prediction items; however, no accuracy differences were found between Repeat and Violation (Table 1). We reasoned that test sensitivity to differences between Strong Repeat vs. Violation items may have been negatively impacted by 1) the change in context between the scanner and the postscan behavioral test; 2) the overall length of the test; and 3) interference from the Weak Prediction items. Changes in context between study and test are known to affect memory performance (Godden and Baddeley 1975), and the large number of items in the postscan test (144 Weak Prediction, 144 Strong Prediction, and 288 foils) may have made the test quite difficult. Thus we conducted a separate behavioral-only experiment using the same paradigm, but with study and test conducted in the same room and testing only Strong Prediction C and D items (with an equal number of foils). As hypothesized, Strong Violation C and D items were better remembered on average than Strong Repeat C and D items, suggesting that prediction violations improved encoding and subsequent memory for Violation items (Table 2). These results support the idea that, when prediction violation is detected in CA1, encoding processes are triggered, i.e., learning by error correction (e.g., Fyhn et al. 2002; Lisman and Grace 2005).

Conclusion

How do we learn from past experience? One method is by using past experience as a model for future experience, i.e., making predictions, assessing the accuracy of those predictions, and using the outcome to improve future predictions. Substructures of the MTL cortex and hippocampus are hypothesized to support prediction generation, detect prediction violations, and consequently encode new information. In this study, we analyzed BOLD activity in the MTL cortex and hippocampus during image sequences with predicted (repeated) and unpredicted (temporally rearranged) orders. Hippocampal subfields CA1 and CA23/DG, as well as MTL cortical regions PRc, were more responsive when images were presented in a rearranged order than when presented in a previously seen order, suggesting that these regions were sensitive to violations of prediction. That is, previous experience with specific presentation orders for specific pictures generated expectations that were then violated (Violation) or confirmed (Repeat). Furthermore, in CA1 there was a significant modulation of the prediction violation response by the level of prediction strength, with stronger predictions (three prior exposures) leading to larger responses than did weaker
predictions (one prior exposure). CA1 consistently contained the greatest proportion of voxels displaying the prediction violation pattern of any ROI. Finally, a follow-up behavioral experiment showed that recognition memory was better for items that had appeared in prediction violation trials than in prediction confirmation trials. These findings provide evidence that CA1 responds when predictions are violated, and that this response is modulated by the strength of predictions based on past experience with specific stimulus orders. These results are compatible with converging evidence and ideas from research in nonhuman animals and anatomically motivated computational models, suggesting that the hippocampus supports the generation of predictions about upcoming events and signals when predictions are violated, via a CA1-supported comparison of expectation and actual outcome. The current study examined neural activity during associative (sequential) learning, providing new high-resolution measurements of hippocampal and MTL cortical dynamics during the acts of prediction, violation detection, and memory encoding, part and parcel of a continuous cycle of events that enables preparation for the future based on past experience.

REFERENCES


