Neurons detect cognitive boundaries to structure episodic memories in humans

Jie Zheng[1], **Andrea G. P. Schjetnan**[2], **Mar Yebra**[3], **Bernard A. Gomes**[3], **Clayton P. Mosher**[3], **Suneil K. Kalia**[2], **Taufik A. Valiante**[2,4,5,6], **Adam N. Mamelak**[3], **Gabriel Kreiman**[1,7,*], **Ueli Rutishauser**[3,8,9,10,*]

[1]Department of Ophthalmology, Children's Hospital, Harvard Medical School, Boston, MA, USA

[2]Krembil Brain Institute and Division of Neurosurgery, University Health Network (UHN), University of Toronto, Toronto, ON, Canada

[3]Department of Neurosurgery, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[4]Department of Surgery (Neurosurgery), Institute of Biomedical Engineering, and Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada

[5]Max Planck-University of Toronto Center for Neural Science and Technology, University of Toronto, Toronto, ON, Canada

[6]Center for Advancing Neurotechnological Innovation to Application, University Health Network, University of Toronto, Toronto, ON, Canada

[7]Center for Brains, Minds and Machines, Cambridge, MA, USA

[8]Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[9]Center for Neural Science and Medicine, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[10]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA

Abstract

While experience is continuous, memories are organized as discrete events. Cognitive boundaries are thought to segment experience and structure memory, but how this process is implemented

remains unclear. We recorded the activity of single neurons in the human medial temporal lobe during the formation and retrieval of memories with complex narratives. Here we show that neurons responded to abstract cognitive boundaries between different episodes. Boundary-induced neural state changes during encoding predicted subsequent recognition accuracy but impaired event order memory, mirroring a fundamental behavioral tradeoff between content and time memory. Furthermore, the neural state following boundaries was reinstated during both successful retrieval and false memories. These findings reveal a neuronal substrate for detecting cognitive boundaries that transform experience into mnemonic episodes and structure mental time travel during retrieval.

Introduction

Our lives unfold over time, weaving rich information into a continuous sequence of experiences. However, our memories are not continuous. Rather, we remember discrete episodes ("events")[1], which serve as anchors to bind together the myriad different aspects (where, when, what) of a given autobiographical memory[2,3], much like objects do in perception[4]. A fundamental unresolved question in human memory is, therefore: what marks the beginning and the end of an episode[5]?

The transformation from ongoing experience to distinct events is thought to rely on the identification of boundaries that separate events[1,6–11]. This theory is motivated by large-scale patterns of activity changes in the human brain around event boundaries[5,12,13], but the underlying neural mechanisms and their relationship to memory are unknown. Neurons in the rodent hippocampus elevate their firing rates in the vicinity of investigator-imposed spatial boundaries[14] and the place fields of hippocampal neurons are shaped by physical boundaries[15][16,17]. In accordance with the boundaries of subenvironments[14], hippocampal place fields remap[18,19] in response to context shifts, and are reinstated[15,20] when the animal is placed back into a familiar context. Additionally, rodent hippocampal neuron ensembles encode lap-specific representations in a maze irrespective of an animal's spatial location[21], presumably representing cognitive boundaries between distinct events. Boundaries shape mnemonic representations of both spatial environments and the events that occur during navigation, and structure the place fields and event-specific representations of cognitive maps. No such understanding at the single-cell level exists for the non-spatial episodic memories that define us as individual human beings[2,22].

We investigated the neuronal mechanisms underlying the identification of event boundaries in humans under semi-realistic continuous experience. We recorded single neuron activity from patients with drug-resistant epilepsy implanted with depth electrodes[23], while testing their memory for the content of video clips with two kinds of embedded cognitive boundaries: soft-and hard boundaries. Soft boundaries are episodic transitions between related events within the same movie, while hard boundaries are episodic transitions between two unrelated movies. Behaviorally, both soft and hard boundaries enhanced recognition of video clip content that followed a boundary, whereas hard boundaries impaired memory of the temporal order between events. We found neurons in the medial temporal lobe that signaled the timing of both types of boundaries. The activity of these

boundary-responsive neurons predicted memory strength as assessed by scene recognition and temporal order discrimination accuracy.

Results

Boundaries boost recognition but disrupt serial order memory

We studied how boundaries influence the formation and retrieval of memories of brief video clips. Twenty patients performed the task while we recorded the activity of single neurons (Fig. 1e; Extended Data Fig. 1, Supplementary Tables 2–3 show the patient demographics and the location of microwire bundles). The task consisted of three parts: encoding, scene recognition, and time discrimination. During encoding (Fig. 1a), subjects watched 90 different and novel video clips containing either no boundaries (NB, one continuous movie shot), soft boundaries (SB, cuts to a new scene within the same movie), or hard boundaries (HB, cuts to a new scene from a different movie; Fig. 1b). A question about the prior movie appeared every 4–8 clips (e.g., is anyone wearing glasses?). Subjects answered $89 \pm 5\%$ of these questions accurately.

We subsequently evaluated what subjects remembered about the video clips with two memory tests: scene recognition (Fig. 1c) and time discrimination (Fig. 1d). During the scene recognition test, subjects were presented with a single static frame. These frames were chosen with equal probability from either the previously presented video clips ("targets"), or from other video clips that were not shown to the subjects ("foils"). Subjects made an "old" or "new" decision together with a confidence rating (sure, less sure and very unsure) in each trial. During the time discrimination test, subjects were shown two old frames chosen from the same video clip, presented side-by-side, and had to indicate which frame was seen earlier in time together with a confidence rating.

In the time discrimination task, subjects correctly identified which frame was shown first in $73 \pm 7\%$ and $73 \pm 8\%$ of trials when the two frames were separated by no boundary (NB) or a soft boundary (SB), respectively (Fig. 2a, green and blue; both above chance of 50%; one sample t-test, NB: $t_{19} = 13.97$, $p = 2\times10^{-11}$; SB: $t_{19} = 11.63$, $p = 4\times10^{-10}$). In contrast, subjects performed significantly worse when discriminating the order of frames separated by a hard boundary (HB) (Fig. 2a, red, HB: $53 \pm 5\%$; $p = 0.02$ against chance level; significantly lower than NB and SB: $F(2, 57) = 51.33$, $p = 2\times10^{-13}$). Subjects also showed longer reaction times (Fig. 2b; HB: 2.10 ± 0.37 seconds; NB: 1.62 ± 0.28 seconds; SB: 1.59 ± 0.34 seconds; $F(2, 57) = 14.25$ $p = 10\times10^{-6}$) and lower confidence ratings when discriminating the order of frames separated by a hard boundary (Fig. 2c; HB: 1.95 ± 0.45; NB: 2.52 ± 0.29; SB: 2.59 ± 0.23; $F(2, 57) = 20.41$, $p = 2\times10^{-7}$). This effect on reaction times and confidence was not driven by accuracy differences as it was observed for both correct and incorrect trials independently (Supplementary Fig. 1). Discriminating the temporal order of two frames was not possible by reasoning alone without having seen the video clips (Supplementary Fig. 2).

Across all trials, the ability to recognize a frame as 'old' did not differ significantly between the types of boundaries preceding the frame (Extended Data Fig. 2a; NB: $76\% \pm 10\%$; SB: $75\% \pm 9\%$; HB: $75\% \pm 8\%$, $F(2, 57) = 0.07$, $p = 0.94$). The reaction times and confidence

ratings during the scene recognition task were also similar across the different types of boundaries (reaction time: Extended Data Fig. 2b; NB = 1.47 ± 0.18 seconds, SB = 1.43 ± 0.16 seconds, HB = 1.49 ± 0.15 seconds, $F(2, 57) = 0.28$ $p = 0.76$; confidence rating: Extended Data Fig. 2c; NB = 2.60 ± 0.18, SB = 2.60 ± 0.20, HB = 2.52 ± 0.28, $F(2, 57) = 0.54$, $p = 0.56$). Therefore, the impaired time discrimination ability across HB transitions was not due to differences in memory strength as measured by scene recognition accuracy. Even though the overall accuracy was similar among NB, SB, and HB conditions, the recognition accuracy of target frames decreased as a function of the time elapsed between the target frame and its immediately preceding boundary. Target frames presented shortly after a SB and HB were remembered better than those farther away from the boundary (Fig. 2e and 2f; SB: $r = -0.61$, $p = 4 \times 10^{-4}$; HB: $r = -0.44$, $p = 0.015$). In contrast, recognition accuracy did not differ significantly as a function of time relative to NBs (Fig. 2d, Pearson correlation; NB: $r = 0.085$, $p = 0.65$). The temporal distance to boundary effects were unidirectional: the temporal distance to future boundaries did not correlate with memory performance (Supplementary Fig. 3a–b). No temporal distance effect was present in the time discrimination task accuracy either (Supplementary Fig. 3c–d). Also, the scene recognition accuracy and time discrimination accuracy were not significantly related to the time at which the tested frames were shown during encoding (Supplementary Fig. 4; scene recognition: $F(5, 114) = 1.87$, $p = 0.11$; time discrimination: $F(5, 114) = 1.1$, $p = 0.37$). Together, the behavioral analyses revealed that frames that closely followed a soft or hard boundary were more likely to be remembered. Temporal order memory, on the other hand, was disrupted by the presence of hard boundaries. These results reveal a tradeoff in the effect of hard boundaries on memories, with enhanced scene recognition memory and disrupted temporal order memory.

Medial temporal lobe neurons demarcate episodic transitions

We next investigated the neuronal responses to boundaries and their relationship to memory by recording from neurons in the medial temporal lobe (MTL; including the hippocampus, amygdala and parahippocampal gyrus; Fig. 1e and Extended Data Fig. 1) and other brain areas (Supplementary Tables 2 and 4). Across all areas, we recorded the activity of 985 neurons from 19 subjects (1 of the 20 subjects yielded no usable recordings, Supplementary Table 2). Of these 985 neurons, 580 were recorded from the MTL. We first tested whether neurons changed their activity following boundaries by comparing their firing rate in a 1s long window following boundaries relative to baseline (1s prior to boundary; Methods). For video clips with no boundaries, we aligned responses to the middle of the clip, and compared responses between before and after this virtual boundary. Figures 3a–b show two example neurons recorded from the hippocampus and parahippocampal gyrus, respectively. These neurons showed a transient increase in firing rates within approximately 300 milliseconds after both soft (blue) and hard (red) boundaries. No such change was observed in the clips without boundaries (green). We refer to this type of neuron as a *boundary cell*. 42/580 MTL neurons (7.24%; expected proportion by chance for all MTL neurons= 2.11%) were classified as boundary cells (Fig. 3c). The regions with the largest proportion of boundary cells were the parahippocampal gyrus (n= 18/68, 26.47%), amygdala (n= 12/169, 7.10%), and the hippocampus (n= 12/343, 3.50%). These proportions are all significantly larger than expected by chance (p < 0.05; Supplementary Table 4).

Was the response of boundary cells a result of the abrupt changes in pixel-level content between the frame before and after the boundary? To answer this question, we considered the responses of the cells during other abrupt changes of visual input: video clip onsets (Fig. 3d) and offsets (Extended Data Fig. 3) during encoding, and image onsets and offsets during scene recognition and time discrimination (Extended Data Fig. 4). Boundary cells did not respond significantly to either clip or image onsets or offsets ($p > 0.05$; permutation t-test, Methods), showing that the response of boundary cells is likely related to higher level cognitive discontinuities rather than pure visual changes.

We also found a second group of neurons that transiently increased their firing rate only following hard, but not soft boundaries or no boundaries. Two examples of such cells, located in the amygdala and hippocampus, are shown in Fig. 3e–f. We refer to this type of neuron as an *event cell*. 36/580 MTL neurons (6.20%; proportion expected by chance for all MTL neurons = 2.26%) were classified as event cells (Fig. 3g). The regions with the largest proportion of event cells were the hippocampus (n= 27/343, 7.87%), amygdala (n= 7/169, 4.27%), and parahippocampal gyrus (n= 2/68, 2.94%). These proportions are all significantly larger than expected by chance (p < 0.05; Supplementary Table 4). Similar to boundary cells, event cells did not significantly change their firing rates following video clip onsets or offsets (Fig. 3h and Extended Data Fig. 3) during encoding, nor did they respond to image onsets or offsets during scene recognition or time discrimination (Extended Data Fig. 4; $p > 0.05$, permutation t-test; Methods). However, boundary cells and event cells did increase their firing rates to the randomly interspersed probe questions that followed some clip offsets (randomly present every 4–8 trials, Supplementary Fig. 6). Seventy six of 580 (13.1%, $p = 0.01$) and four of 580 (0.7%, $p = 0.43$) MTL cells changed their firing rate in response to clip onsets and clip offsets, respectively, but neither of these cells qualified as boundary cells or event cells (Extended Data Fig. 5).

Soft and hard boundaries differed in terms of their high-level conceptual narrative, which is interrupted in HBs but not in SBs. To evaluate whether it is possible to determine from visual features alone whether a boundary is soft or hard, we computed the differences between pre-and post-boundary frames in terms of pixel-level characteristics (i.e., luminance, contrast, complexity, entropy, color distribution), high-level visual features (i.e., objects), and perceptual similarity ratings. These analyses revealed that SB and HB transitions did not differ significantly from each other in any of the attributes we tested (Supplementary Table 1). Therefore, the differential activation of event cells to HBs but not SBs was likely a result of detection of the disruption in the conceptual narrative, that is, a transition between two different episodes.

While both boundary cells and event cells responded to HB transitions, a comparison of their response dynamics indicated that boundary cells responded to hard boundaries approximately 100 milliseconds earlier than event cells (Fig. 3i). This latency difference was also observed when comparing the time at which the peak responses were reached: boundary cells showed a peak at 197 ± 49 milliseconds, whereas event cells showed a peak at 301 ± 55 milliseconds (Fig. 3j, $F(1, 76) = 274.78$, $p = 6\times10^{-27}$).

We also evaluated the existence of boundary and event cells in brain areas other than the MTL, such as the medial frontal cortex, insula, and orbitofrontal cortex (OFC). We found 8/405 (1.96%) boundary cells and 9/405 (2.22%) event cells among the non-MTL group (Supplementary Tables 2 and 4), with only event cells in the OFC exceeding the proportions expected by chance. Thus, boundary responsive neurons are largely within the MTL, where we restricted the following analyses.

Responses of boundary and event cells predict memory strength

We next asked whether the responses of boundary cells and event cells during encoding correlated with later measures of memory for the content of the videos. We examined whether the strength of responses of boundary cells or event cells to boundaries varied as a function of whether the familiarity or temporal order of a stimulus was later remembered or forgotten. Fig. 4a shows an example boundary cell located in the hippocampus whose response during encoding differed between video clips from which frames were later correctly remembered as familiar (Fig. $4a_1$) vs. incorrectly identified as novel (forgotten, Fig. $4a_2$): the responses to boundaries that preceded later remembered frames were stronger. This effect was present, on average, among boundary cells (n = 42) for frames preceded or followed by both SBs and HBs, but not by NBs (Fig. 4c; SB: $F(1, 82) = 82.93$, $p = 4 \times 10^{-14}$; HB: $F(1, 82) = 156.9$, $p = 1 \times 10^{-20}$; NB: $F(1, 82) = 1.18$, $p = 0.28$; Supplementary Fig. 8, SB: $F(1, 82) = 91.67$, $p = 5 \times 10^{-15}$; HB: $F(1, 82) = 62.78$, $p = 1 \times 10^{-11}$; NB: $F(1, 82) = 0.05$, $p = 0.83$). This effect was specific to scene recognition and boundary cells. First, the firing rate of boundary cells did not significantly predict performance in the time discrimination task (Extended Data Fig. 6a and Extended Data Fig. 6c; NB: $F(1, 82) = 1.25$, $p = 0.27$; SB: $F(1, 82) = 1.35$, $p = 0.25$; HB: $F(1, 82) = 1.14$, $p = 0.29$). Second, the firing rate of event cells (n = 36) during encoding was not predictive of scene recognition memory (Extended Data Fig. 7a and Extended Data Fig. 7c; NB: $F(1, 70) = 1.12$, $p = 0.29$; SB: $F(1, 70) = 1.63$, $p = 0.21$; HB: $F(1, 70) = 0.79$, $p = 0.38$).

Given the importance of theta-frequency band spike field coherence in plasticity[24], we next considered the timing of spikes with respect to the theta band in the local field potentials (LFP, 4–8Hz, measured on the same microwire from which the neuron was recorded from, Methods; see Supplementary Fig. 9 for power spectra). We determined the theta phase of each spike that occurred within 1s following boundaries and compared the resulting phase distributions among NB, SB and HB. Event cells tended to fire at a consistent phase of the theta band LFP following both HB and SB boundaries for clips whose temporal order was later remembered correctly (Fig. 4f shows an example). To summarize this effect across the population, we computed the mean resultant length (MRL) across all phases for all spikes fired by a given cell (Methods). If the time of spikes are randomly distributed, the MRL equals 0, whereas an identical phase for all spikes would result in an MRL of 1. The mean MRL across all event cells (n = 36) was significantly larger following both SB and HB but not NB if temporal order was later correctly remembered (Fig. 4h; SB: $F(1, 70) = 81.55$, $p = 2 \times 10^{-13}$; HB: $F(1, 70) = 60.79$, $p = 4 \times 10^{-11}$; NB: $F(1, 70) = 1.53$, $p = 0.22$). This effect was specific to event cells and temporal order memory. First, the strength of theta phase-locking of event cells did not predict scene recognition memory success (Extended Data Fig. 7b and Extended Data Fig. 7d; NB: $F(1, 70) = 0.75$, $p = 0.39$; SB: $F(1, 70) = 1.1$,

$p = 0.30$; HB: $F(1, 70) = 2.13$, $p = 0.15$). Second, the strength of phase-locking of boundary cells neither predicted scene recognition memory success (Fig. 4b,d; NB: $F(1, 82) = 1.16$, $p = 0.28$; SB: $F(1, 82) = 1.87$, $p = 0.18$; HB: $F(1, 82) = 0.45$, $p = 0.5$) nor temporal order memory (Extended Data Fig. 6b and Extended Data Fig. 6d; $F(1, 82) = 1.33$, $p = 0.25$; SB: $F(1, 82) = 0.14$, $p = 0.71$; HB: $F(1, 82) = 1.98$, $p = 0.16$). Third, we evaluated whether there were cells whose theta-band phase locking of spikes following boundaries was predictive of the success of memory formation regardless of whether their firing rate was modulated (Methods). There were 32 out of 580 MTL cells that showed an enhanced number of spikes phase-locked in the theta band after a boundary compared to before a boundary and where the phase-locking was correlated with correct/incorrect performance in either one of the two memory tasks. Out of those 32 cells, 20 (56%) were also event cells. In contrast, there was no significant overlap between the 32 cells and boundary cells (Supplementary Table 5).

In summary, boundary cells and event cells predicted distinct aspects of memory formation: whereas the firing rate of boundary cells was predictive of later scene recognition memory performance, the phase-locking of event cells was predictive of temporal order memory performance.

Neural state shifts across boundaries reflect memory strength

We next investigated the changes in the neural responses following boundaries at the population level of all n=580 MTL cells (pseudopopulation, Methods). We examined the dynamics of population activity around the boundaries by evaluating the change of activity using principal component analysis. During NB video clips, the neural state exhibited only slow changes as a function of time (Fig. 5a, black dot marks the middle of the clip). In contrast, the neural state changed abruptly following SBs and HBs (Fig. 5b–c; black dot marks boundary time). These abrupt 'neural state shifts' were consistent with the changes in firing rates we reported for boundary cells and event cells, but the observations in Fig. 5 reflect the activity of all MTL cells. To quantify the size of state shifts, we computed the multidimensional Euclidean distance MDD(t) in state space between a given time t and the boundary (Fig. 5d–g). The dimensionality of the state space we used was the number of PCs that explained 99% of the variance. Plotting MDD as a function of time revealed an abrupt change within ~300 ms after the boundary for SB and HB video clips (Fig. 5d–g). This abrupt change can also be seen at the level of individual subjects (Extended Data Fig. 8).

We evaluated what types of cells contributed most to the neural state shift. First, neural state shifts following SBs were only visible when boundary cells were included (Fig. 5d–e). Second, early neural state shifts after HBs were only visible when event cells were included while later HB-related shifts remained present in the absence of either event cells (Fig. 5f) or both event and boundary cells (Fig. 5g). Third, the point of time at which MDD reached its maximal value varied systematically between groups of cells: the responses carried by boundary cells appeared significantly earlier than those carried by event cells and non-boundary/event cells (Fig. 5h; $F(3, 76) = 103.96$, $p = 8 \times 10^{-27}$). Together, this shows that early population-level state shifts are principally due to the activity of boundary

cells, whereas event cells and non-boundary/event cells in MTL contribute to slower-latency HB-related state shifts.

We next assessed whether the size of neural state shifts following boundaries during encoding was related to whether a stimulus was later remembered or not. We computed the extent of state changes in the population following a boundary by calculating the total Euclidean distance traversed in state space in the 0–1s time window after boundaries (AUC MDD; Methods). The AUC MDD was positively correlated with recognition accuracy for frames following SBs and HBs, but not NBs (Fig. 5i; SB: $r = 0.653$, $p = 0.002$; HB: $r = 0.565$, $p = 0.009$; NB: $r = 0.214$, $p = 0.256$). In contrast, the AUC MDD was negatively correlated with accuracy in the time discrimination task for HBs, but not for NBs or SBs (Fig. 5j, HB: $r = -0.677$, $p = 0.001$; NB: $r = 0.212$, $p = 0.261$; SB: $r = -0.273$, $p = 0.244$). Together, these results reveal a neural correlate of the tradeoff between these two types of memory, with large neural state shifts beneficial for scene recognition memory but detrimental for order memory.

Neural context after boundaries reinstates during recognition

It is thought that reinstatement of the neural context present at encoding enables mental time travel during memory retrieval[25,26]. However, it remains unknown what exactly is reinstated for continuous experience and how boundaries shape the retrieval process. To address this question, for each subject, we quantified the degree and timing of reinstatement by computing the correlation between the vectors of spike counts of all recorded MTL neurons during the scene recognition task (1.5s fixed time window) and during encoding (1.5s sliding window, step size 0.1s; Methods). Correct targets (i.e., frames from presented clips during encoding that were correctly remembered as "old") were accompanied by a significant positive correlation between neural activity during the scene recognition and the encoding period shortly after SB/HB transitions (Fig. 6a,e; $p < 0.01$, permutation test, Methods). In contrast, we observed no significant correlation for forgotten targets (i.e., frames from presented clips that were incorrectly marked as "new"; Fig. 6b,f). This effect could not be explained by subjects not attending to the scene recognition task because visually responsive cells responded equally well to both remembered and forgotten target trials (Extended Data Fig. 9).

The reinstated neural context during retrieval was most similar to the neural context present during encoding approximately ~1.2s after the boundary that preceded the tested frame (Fig. 6i, filled circles; SB: -1.26 ± 0.38 seconds, $t_{19} = 14.68$, $p = 8 \times 10^{-12}$; HB: -1.28 ± 0.48 seconds, $t_{19} = 11.80$, $p = 3 \times 10^{-10}$). Notably, the point of time of maximal similarity preceded the time at which the later tested frame was shown by 1–1.5s (Fig. 6i, empty circles; SB: 1.53 ± 0.61 seconds, $t_{19} = 11.18$, $p = 8 \times 10^{-10}$; HB: 1.72 ± 1.03 seconds, $t_{19} = 7.44$, $p = 5 \times 10^{-7}$). This observation remains true also for smaller window sizes used to compute the correlations (Supplementary Fig. 13), indicating that the neural state reinstated during testing is the one that was present in between the preceding boundary and the tested frame. Thus, the neural state that was reinstated is the one present well before the tested frame was shown during encoding. Together with an absence of significant reinstatement during incorrect targets (Fig. 6B), these analyses suggest that the neural

correlate of reinstatement is not the result of identical sensory inputs in the clips and tested frames. Besides, no significant correlation for correctly identified foils was observed (i.e., frames from unpresented clips correctly marked as "new"; Fig. 6c and 6g).

Reinstatement is thought to contribute primarily to the recollection of memories that are remembered with high confidence[27,28]. Compatible with this view, the correlations following boundary transitions were significantly stronger in high- compared to low-confidence trials (converted from three-levels of confidence; Methods). This difference was observed during both correctly remembered targets and falsely recognized foils (Fig. 6j–m; SB, correct targets: $p = 5\times10^{-10}$; HB, correct targets: $p = 4\times10^{-6}$; NB, correct targets: $p = 0.79$; SB, incorrect foils: $p = 5\times10^{-7}$; HB, incorrect foils: $p = 5\times10^{-5}$; NB, incorrect foils: $p =0.18$; t-test).

Notably, strong correlations between the neural state at retrieval and encoding also occurred when a new image was incorrectly classified as seen before (Fig. 6d and 6h; incorrect foil; $p < 0.01$, permutation test; Methods), thereby revealing a neural explanation for the false alarms. Were these false alarms, which were accompanied by neural reinstatement, caused by visual similarity between the targets and foils? To address this question, we assessed the similarity between each target and foil by acquiring similarity ratings from an independent control group of subjects (n = 30). Similarity values were balanced across NB, SB and HB (Supplementary Fig. 2a). We split foils into low (top 66.67% – 100%), medium (top 33.33% – 66.67%) and high (top 1% – 33.33%) similarity group. Correlations between encoding and scene recognition were significantly stronger for highly similar foils falsely recognized as old compared to low and medium similarity foils (Fig. 6o; incorrect foil: $F(2, 54) = 10.67$, $p = 1\times10^{-4}$). In contrast, the extent of correlation for correctly rejected foils (true negatives) did not vary as a function of similarity (Fig. 6n, correct foil: $F(2, 54) = 2.182$, $p = 0.144$). Thus, the reason for false alarms is that the wrong context is reinstated due to the high similarity of the foil with a target. Together, these results support the notion that the neural state present at encoding following the boundary is reinstated during memory retrieval.

Discussion

Memories are often conceptualized as discrete events on a narrative timeline[6]. However, the very definition of an event remains enigmatic. Where do events start and end, and how are multiple signals bound together over time to form a singular event? Here, we test the hypothesis that boundary detection is a mechanism that segments continuous experience into discrete events. Behaviorally[29], boundaries enhance scene recognition memory for temporally proximal events while disrupting temporal order memory. We found two types of cells that responded to cognitive boundaries: one responded to both soft and hard boundaries, while another group responded only after hard boundaries.

Both soft-and hard boundaries are accompanied by salient visual changes, whereas the 'no boundary' control condition does not (Fig. 1b). However, the observed responses to boundaries (Fig. 3) cannot be explained by these sharp visual input changes. First, cells differentiate between soft-and hard boundaries even though both types encompass similar changes at the pixel level (Supplementary Table 1). Second, boundary and event cells did not

respond to strong visual changes at clip onsets (Fig. 3d,h) or offsets (Extended Data Fig. 3) during encoding, nor to the image onsets or offsets during the scene recognition and time discrimination tasks (Extended Data Fig. 4). In our task, trial structure is predictable, i.e., the inter-trial interval is always followed by a video and then the next inter-trial interval (Fig. 1a–d). In contrast, whether a boundary will occur in a given video and if so what type (SB or HB) is not predictable. One hypothesis for the absence of responses to clip onsets and offsets is therefore their predictability, which would lead to an absence of prediction errors that are hypothesized to underly event segmentation[30]. Supporting this hypothesis, boundary and event cells also increased their responses following the unpredictable probe question during encoding (present randomly every 4–8 trials; Fig. 1a). In contrast to the selective single neuron response to cognitive boundaries, prior fMRI work[31] reported a clip-offset triggered BOLD signal change in the hippocampus whose magnitude was predictive of subsequent memory strength. This off-response has been interpreted to have the same origin as the between-event responses[32] and was present despite the fixed clip length and trial structure.

Boundary cells respond to both soft-and hard boundaries, whereas event cells respond only to hard boundaries (Fig. 3). These distinct responses may reflect the hierarchical structure of episodic memory, with event cells representing episodic transitions between distinct events while boundary cells represent more frequent but smaller episodic transitions within the same overall narrative. These findings provide evidence for the theory that event segmentation is a hierarchical dynamic procedure, with fine to coarse segmentations associated with different kinds of cognitive boundaries[9]. The anatomical location and response latency of the cells are also compatible with this proposal: boundary cells respond first and are most common in the parahippocampal gyrus, whereas event cells respond later and are most common in the hippocampus (Supplementary Table 4). This distinction is also visible at the population level, with neural state shifts for SBs mainly driven by boundary cells, whereas HB-related state shifts occur later and are driven by a broader group of cells (Fig. 5d–h). Notably, the late HB-related state shifts are partially driven by cells that are not classified as event cells (a conclusion that holds even when using a more liberal definition of event cells at p<0.01). This suggests that besides the early HB-related responses, there is a secondary later response to HBs that is encoded as a distributed population response. We hypothesize that the early responses of boundary cells reflect contextual changes detected in the higher-level visual areas[33,34], while event cells are the result of a late output signal from a comparator operation between predicted and received signals[35,36].

Responses of boundary or event cells bring to mind border and place cells in the rodent hippocampus[16,37]. As rodents move between compartments, place cells cluster at boundaries (e.g., doorways)[14], crossing of which is followed by remapping[18,19] or reinstatement[15,20] of a different set of hippocampal place fields. Here, boundary cells and event cells respond to transitions (boundaries) between different episodic contexts. Also, similar to place field remapping, the neural state changed abruptly following a boundary. When subjects are re-exposed to familiar target frames during the later recognition test, the neural state reinstates if the item is successfully recognized. Similar to place field reinstatement, the reinstated neural state is most similar to the one following boundaries even before when the tested frame is shown. This finding provides insight into the question of what neural context is reinstated during mental time travel and memory search[38–42]. This finding also indicates

that abrupt changes in neural context are important to demarcate periods of time that can be reinstated later. We note several differences between boundary and event cells and border cells in rodents. Border cells respond to physical boundaries and are observed in tasks in which rodents are extensively trained. In contrast, boundary or event cells in humans respond to cognitive boundaries in a variety of different videos, none of which subjects have seen before. This property is an essential requirement for a process to divide experience into episodes to shape episodic memories, which, by definition, occur only once in novel environments.

What roles do boundary responses play in episodic memory? At the single-cell level, the firing rate of boundary cells predicts scene recognition memory strength and the clustering of event cells' spike timing relative to theta phase predicts temporal order memory encoding success. This indicates that these two kinds of cells play distinct roles during encoding, with each strengthening only one kind of memory using a different mechanism. The response of boundary and event cells during encoding was "content-invariant" because they responded to many clips with varying content (Fig. 3). The role of boundary responses during retrieval was in guiding the points of time that would later be reinstated (Fig. 6i), but not participating in the reinstatement process itself. This is expected if boundary-cells do not carry information about content. Confirming this, the results on reinstatement remain essentially unchanged after excluding boundary and event cells from the analysis (Supplementary Fig. 12). Together, these findings suggest that boundary and event cells play two roles in episodic memory: First, they structure memories during encoding, and they serve as markers for periods of time that are later reinstated.

Methods

Subjects

Patients: Twenty patients with drug-resistant epilepsy volunteered for this study and gave their informed consents. The institutional review boards of Toronto Western Hospital and Cedars-Sinai Medical Center approved all protocols. The task was conducted while the patients stayed in the hospital after implantation of depth electrodes for monitoring seizures. The location of the implanted electrodes was solely determined by clinical needs. The behavioral analyses included results from all 20 subjects and the neural results were analyzed across 19 subjects (Subject #20 had no usable recordings, see patient demographics in Supplementary Table 2).

Amazon Mechanical Turk Workers (MTurk workers): MTurk workers were recruited for similarity ratings (see Methods section Similarity ratings), including 30 subjects (age 23.25 ± 3.42, 9 female) for rating the visual properties of different boundaries (Supplementary Table 1), 30 subjects (age 22.79 ± 5.73, 11 female) for rating the similarity between target and foil frames (Supplementary Fig 2a), and 30 subjects (age 25.06 ± 6.11, 7 female) for performing the time discrimination task without encoding session (Supplementary Fig. 2b). All control tasks conducted on Amazon Mechanical Turk were under the approval of the institutional review board of Boston Children's Hospital and informed consents were obtained with digital signatures for each subject.

Task

The task consisted of three parts: encoding, scene recognition, and time discrimination (Fig. 1a–d).

Encoding: subjects watched a series of 90 unique clips with no sound and were instructed to remember as much of the clips as possible. Each trial started with a baseline period: a fixation cross reminding subjects to fixate at the center of the screen throughout the task. The duration of the baseline period ranged from 0.9s to 1.1s (randomized, sampled from uniform distribution). The fixation period was followed by the presentation of a video clip that contained either no boundaries (NB, continuous movie shots; virtual 'NB' boundaries for analysis purposes were always located in the middle of the clip), soft boundaries (SB, cuts to a new scene within the same movie, 1 to 3 SB per clip, randomly distributed in the clips), or a hard boundary (HB, cuts to a new scene from a different movie, 1 HB per clip located at 4 seconds after the start of the clip). Examples of SB and HB are shown in Fig. 1b. A yes/no question related to the content of the clip (e.g., is anyone in the clip wearing glasses?) appeared randomly after every 4–8 clips.

Scene recognition: After watching all 90 clips, subjects' memory for the content of the videos was evaluated in a scene recognition test. During scene recognition, frames extracted from encoded clips (target frames), and frames from new, never shown, clips (foil frames) were presented to the subjects. Subjects were instructed to identify whether these frames were "old" or "new" (i.e., whether they had seen the frame during the encoding session). To generate the testing frames for scene recognition, we first extracted two frames from each clip, one randomly pulled out from the first half of the clip and the other one from the second half. We then kept half of these frames extracted from first/second half of the clip (in total n = 90) as target frames and used the other half as templates to search for foil frames (in total n = 90) from a different movie played by different actors/actresses (n = 30), a different movie played by the same actors/actresses (n = 30), or the unpresented portion from the same movie played by same actors/actress (n = 30) to introduce different levels of similarity between the target frames and foil frames. The total number of target/foil frames (30 target frames and 30 foils frames for each boundary type) and the average similarity level of foil frames were counterbalanced across different boundary types ($F (2, 87) = 1.72$, $p = 0.19$; rated by Amazon Mechanical Turk workers, see Methods section Similarity ratings).

Time discrimination: After the scene recognition test, we evaluated subjects' memory about the temporal structure of the clip with a time discrimination test. In each trial, two frames (half of them picked at 1s and 7s, and the other half picked at 3s and 5s of the clip) separated by different kinds of boundaries (NB, SB, or HB) were extracted from the same video clip and were presented side by side. Subjects were instructed to indicate which of the two frames (i.e., "left" or "right") appeared first (earlier in time) in the videos they watched during encoding. In both the time discrimination and recognition memory test, the duration of the baseline period ranged from 0.45–0.55s (randomized, sampled from uniform distribution).

Confidence measurement: All binary choices through the encoding session, scene recognition and time discrimination were made together with a subjective confidence judgment (i.e., sure, less sure, very unsure). Thus, there were always 6 possible responses for each question. Given that there were few "less sure" and "very unsure" responses compared to "sure" responses, we grouped "very unsure" and "less sure" responses together as "low confidence" and labeled "sure" responses as "high confidence" in Fig. 6j–m.

Similarity ratings

Visual properties of SB and HB.—Both SB and HB transitions were accompanied by transient visual changes (cuts to a new scene), whereas there were no such visual changes for the NB condition. We quantified the visual changes of each boundary type by computing metrics that relate to pixel level differences: Luminance; Contrast; Complexity; Entropy; Color distribution) between pre- and post-boundary frames. In addition, to quantify visual differences not directly captured at the pixel level, we used pre- and post-boundary frames as inputs for the AlexNet network (pretrained on ImageNet dataset)[43], extracted the activation matrices from the layer 'fc7' for both images and computed the Euclidean distance between their activation matrices. Moreover, we collected perceptual ratings (i.e., similarity ratings between pre- and post- boundary frames) from Amazon Mechanical Turk (MTurk) workers. During similarity ratings, pre- and post-boundary frames were presented side by side and MTurk workers were instructed to rate the similarity of the image pairs by clicking on a scaling bar ([0 1]; 0 = different, 1 = identical). See results in Supplementary Table 1.

Similarity ratings between target and foil frames: When selecting foil frames, we used target frames as templates to search for foil frames with different similarity levels (see Methods section Task). We presented the target frame with its corresponding foil frame side by side and instructed MTurk workers to rate the similarity between them (see results in Supplementary Fig. 2a).

Time discrimination without encoding.—To ensure the time discrimination task could not be solved by pure reasoning, we recruited MTurk workers to perform the time discrimination test without watching clips (Supplementary Fig. 2b).

Electrophysiology

We recorded bilaterally from the amygdala, hippocampus, and parahippocampal gyrus, and other regions outside the medial temporal lobe using hybrid depth electrodes (Ad-Tech company, Oak Creek, Wisconsin, USA), which contained eight 40-µm diameter microwires at the tip of each electrode shank. For each microwire, broadband signals (0.1–9000 Hz filtered) were recorded at 32 kHz using the ATLAS system (Neuralynx Inc., Bozeman, Montana, USA).

Spike sorting and quality metrics of single units

The recorded signals were filtered offline in the 300 to 3000 Hz band, with a zero-phase lag filter. Spikes were detected and sorted using the semiautomated template-matching algorithm Osort[44,45] v4. We computed spike sorting quality metrics for all putative single units (Supplementary Fig. 5) to quantify our recording and sorting quality[46–48].

Electrode localization

Electrode localization was based on postoperative MRI/CT scans. We co-registered postoperative and preoperative MRIs using Freesurfer's mri_robust_register[49]. To summarize electrode positions and to provide across-study comparability, we aligned each subject's preoperative scan to the CIT168 template brain in MNI152 coordinates[50] using a concatenation of an affine transformation followed by a symmetric image normalization (SyN) diffeomorphic transform[51]. The MNI coordinates of the 8 microwires from the same electrode shank were marked as one location. MNI coordinates of microwires with putative neurons detected across all the subjects were plotted on a template brain for illustration (Fig. 1e).

Data analyses

Boundary cell: For each recorded neuron, we counted spikes within the [0 1] seconds (post boundary) and [−1 0] seconds time interval (baseline) relative to boundaries during encoding. A cell was considered a boundary cell if it met the following criteria: (1) its spike counts within post boundary time windows were significantly different from its spike count within baseline time windows for SB and HB but not for NB ($p < 0.05$, permutation t-test); (2) its spike counts within post boundary time windows were significantly greater in SB and HB than NB ($p < 0.05$, permutation t-test).

Event cell: A cell was considered as an event cell if it met the following criteria: (1) its spike counts within post boundary time windows were significantly different from its spike count within baseline time windows for HB but not for NB and SB ($p < 0.05$, permutation t-test); (2) its spike counts within post boundary time windows were significantly greater in HB than NB and SB ($p < 0.05$, permutation t-test).

Boundary and event cells' responses to clip onsets and offsets: For each selected boundary cell and event cells, we counted spikes within the [0 1]s (post) and [−1 0]s (pre) time interval relative to clip onsets/offsets during encoding for clips with NB, SB and HB separately. The boundary cell or event cell was considered as not responding to clip onsets/offsets if their spike counts within each boundary condition did not differ between post and pre window ($p > 0.05$, permutation t-test).

Phase-tuning cells: we computed the mean resultant length (MRL) for the theta-band phases of all spikes fired within a 0–1s window following a boundary and −1–0s preceding a boundary (baseline) during encoding. We also computed the MRLs in the 0–1s post boundary time window separately for spikes fired in trials that were later remembered (correct) vs forgotten (incorrect) during the scene recognition and time discrimination task. We defined a cell as a *phase-tuning neuron* if it met the following criteria: (a) its MRL within the post boundary time window was significantly different from its MRL within the baseline time window for SB and HB but not for NB trials ($p < 0.05$, permutation t-test); (b) its MRL within the post boundary time window was significantly different between correct and incorrect trials in either the scene recognition and/or the time discrimination task ($p < 0.05$, permutation t-test).

Chance level for cell response analyses

To estimate the number of neurons that would be considered boundary cells or event cells by chance, we repeated the same procedures for boundary cell and event cell analyses after randomly shuffling the trial labels (NB, SB, HB) 1000 times. For each iteration, we obtained the proportion of selected boundary cells and event cells relative to the total number of neurons within each region. These 1000 values formed the empirically estimated null distribution for the proportion of boundary cells and event cells expected by chance. A region was considered to have a significant amount of boundary cells or event cells if its actual fraction of significant cells exceeded 95% of the null distribution (Supplementary Table 4; $p < 0.05$).

Association between spiking activity during encoding and later memory retrieval accuracy

Firing rate modulation: For each boundary cell and event cell, we grouped its spike activity within [0 1] seconds after boundaries during encoding based on subjects' subsequent memory performance either in the scene recognition task (correct versus incorrect recognition) or the time discrimination task (correct versus incorrect discrimination). We then computed the firing rate as a function of time (bin size = 200ms, step size = 2ms) for each trial, which was further z-score normalized using the mean and standard deviation of the firing rate across the whole trial. For each boundary cell and event cell, we then computed the mean z-scored firing rate within [0 1] seconds time interval relative to boundaries for each trial and averaged this value across trials within each boundary type. The resulting values across all boundary cells and event cells were used for comparisons across NB, SB and HB conditions (Fig. 4c, Fig. 4g, Extended Data Fig. 6c and Extended Data Fig. 7c).

Phase modulation: For each spike of each boundary cell and event cell, we computed the phase of the spike relative to the theta-frequency band filtered local field potential signals (LFP) recorded from the same microwire. To eliminate potential contamination by the spike waveform, we removed the LFP signal within the 3ms around each spike and replaced it with a linear interpolation. The cleaned LFP signals were then band-pass filtered between 1–300Hz (a zero phase delay finite impulse response filter with Hamming window) and downsampled from 32Khz to 500Hz. We performed automatic artifact rejection[52] and manual visual inspection (using ft_databrowser.m from Fieldtrip toolbox[53] version 20190527) to remove large transient signal changes from the downsampled LFPs. Next, we extracted neural activity within the theta band by bandpass filtering in the 4–8Hz band (eegfilt.m function in EEGLAB toolbox[54]: a two-way, zero phase lag, least-squares finite impulse response filter to prevent phase distortion), followed by the Hilbert transform to obtain theta phase as a function of time. We then extracted the phase for each spike (i.e., spike phases) by boundary cells and event cells. The phase locking strength of each boundary cell or event cell was quantified as the mean resultant length (MRL) of all spike phases of all spikes that occurred within a 0–1s window after boundaries (0 = no phase locking; 1 = strongest phase locking). The resulting MRL values were then compared between trials with correct or incorrect subsequent memory performance for NB, SB and HB trials separately (Fig. 4d, Fig. 4h, Extended Data Fig. 6d and Extended Data Fig. 7d).

The computation of MRL is sensitive to the number of spikes. Therefore, the comparison of MRL between correct and incorrect trials was conducted with balanced spike counts.

State-space analyses

Neural state trajectories: For each trial, we binned each neuron's spike counts during encoding into non-overlapping 10-ms wide bins, followed by smoothing with a 200ms standard deviation Gaussian kernel and z-score normalization (mean and standard deviation were calculated across the entire trial). We used these z-scored smoothed spike density estimates from all recorded MTL cells across all subjects to form a pseudopopulation. We applied principal component analysis (PCA) to reduce the dimensionality of the pseudopopulation (MATLAB R2019b function svd.m). We then rank-ordered the resulting principal components (PCs) by their explained variance (function dpca_explainedVariance.m from dpca toolbox[55]) and plotted the average neural state trajectories for each boundary type in a three-dimensional space constructed by the top three PC components (Fig. 5a–c).

Multidimensional distance (MDD): MDD was defined as the Euclidean distance between two points in the PC space (with all the first n PCs that accounted for > 99% explained variance). Note that while this space captured 99% of explained variance it was nevertheless substantially lower dimensional than the original space.

Area-under-the-curve Multidimensional distance (AUC MDD): AUC MDD was defined as the cumulative sum of all Euclidean distance values within the [0, 1] seconds time window after boundaries (in the PC space).

Reinstatement of neural context

This analysis was done separately for each session of simultaneously recorded neurons and did not rely on the pseudopopulations defined in the previous section.

Correlation between encoding and retrieval: Neural activity was quantified for each neuron in bins of 1.5s width and a step size of 100ms. We computed the Pearson correlation coefficients (corrcoef.m from MATLAB R2019b) between the neural population activity during scene recognition (1 time bin × number of neurons) and encoding (80 time bins × number of neurons) at each time step.

Significant correlation threshold: We computed the same correlation values after randomly shuffling the trial labels (i.e., disrupting the correspondence between encoding and scene recognition trials) to obtain the average correlation strength across trials and neurons expected by chance. This procedure was repeated 1000 times to form a null distribution, in which the 2.5^{th} and 97.5^{th} percentile values were used as the threshold to determine significance of the actual correlation values (dashed horizontal lines in Fig. 6).

Statistics & Reproducibility

No statistical method was used to predetermine sample size, but our sample sizes are similar to those reported in previous publications[46,56]. Data collection and analysis were not performed blind to the conditions of the experiments. The experiments were not randomized.

No data were excluded from the analyses. For comparisons between two groups, we used the permutation t-test statistic, and for comparisons between more than two groups, we used parametric one-way ANOVA. For statistical thresholding, permutation tests were conducted to generate a null distribution estimated from 1000 runs on data with scrambled labels, which avoids the assumption of normality when evaluating significance.
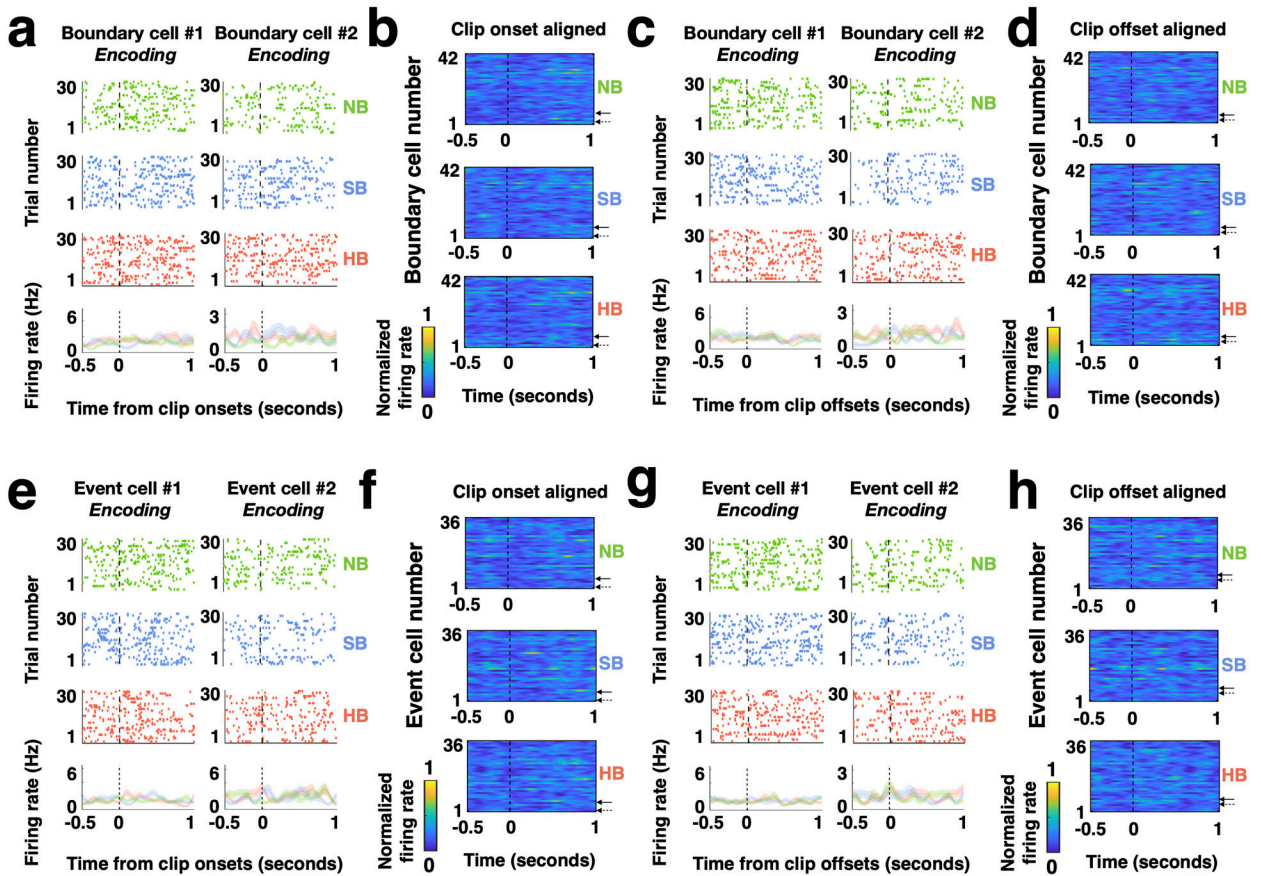
Extended Data
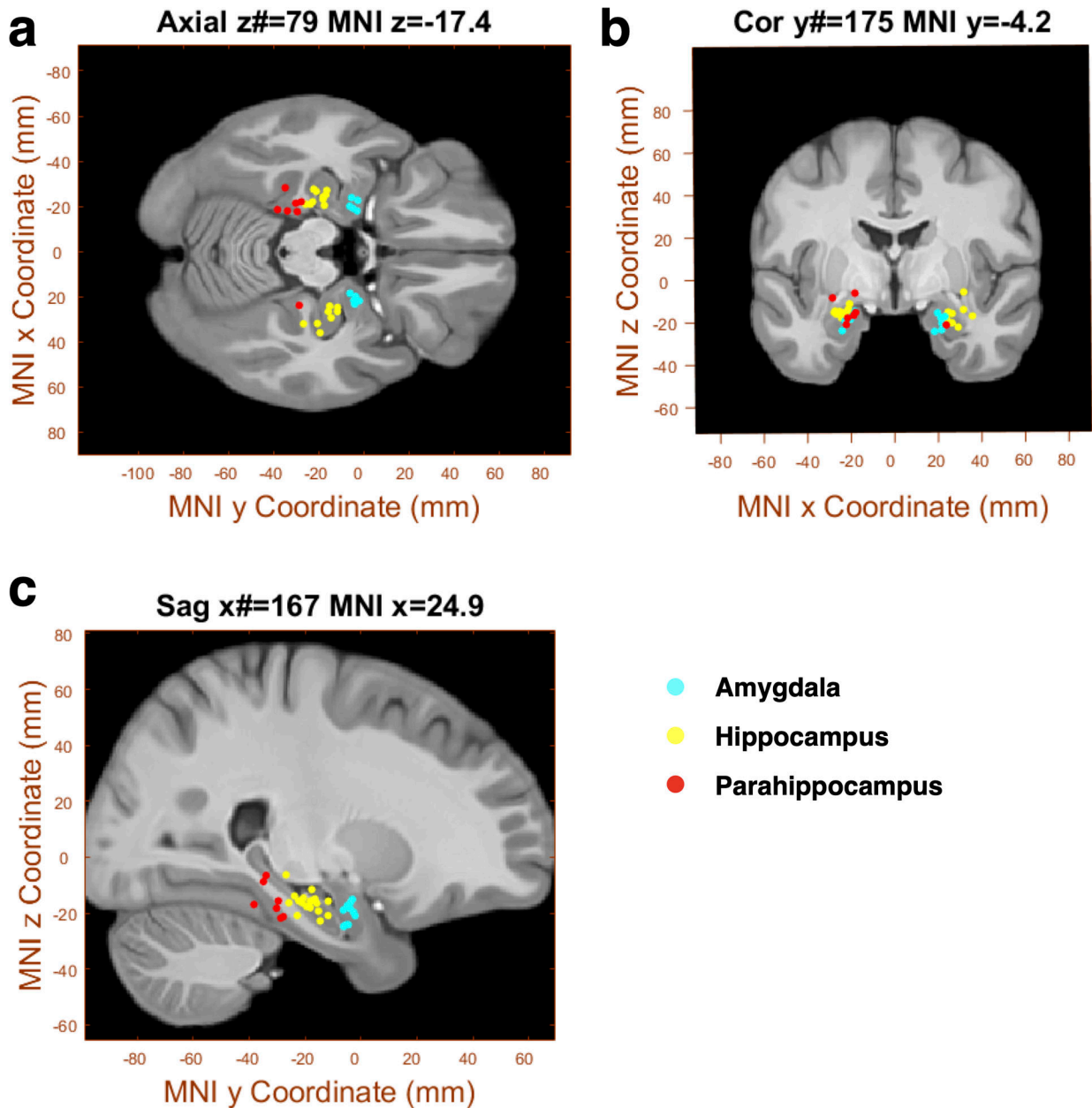
Extended Data Fig. 1. Electrode locations in MNI coordinates, Related to Fig. 1
a-c, Each dot is the location of a microwire bundle in either the amygdala (cyan), hippocampus (yellow) or parahippocampus (red) on which at least one event or boundary

cell was recorded, also presented in a template brain in Fig. 1e. Coordinates are in Montreal Neurological Institute (MNI) 152 space, here plotted on top of the CIT168 brain template for axial (**a**), coronal (**b**), and sagittal (**c**) view (see Methods).
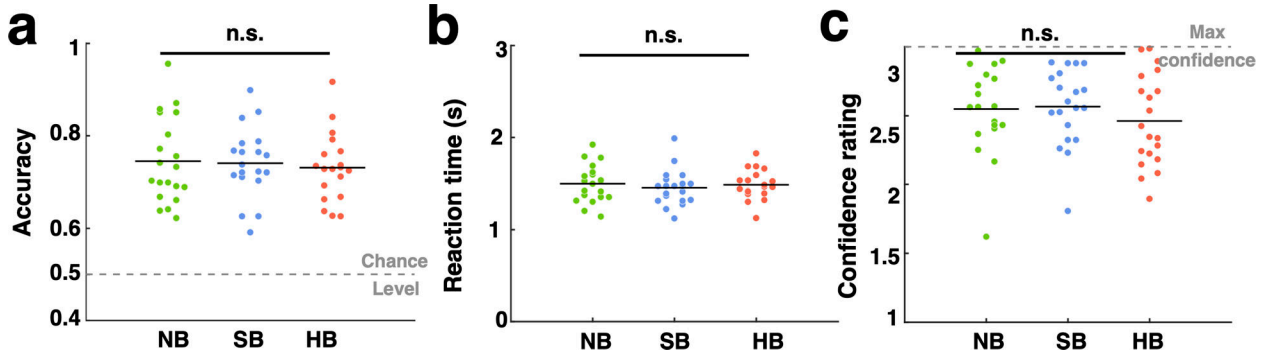

Extended Data Fig. 2. Subjects' performance in the scene recognition task did not differ significantly across different boundary types, Related to Fig. 2

a-c, Behavior quantified by accuracy (**a**), reaction time (**b**), and confidence level (**c**) across all trials. Results are shown for boundary type NB (green), SB (blue), and HB (red) during the scene recognition task. The horizontal dashed lines in (**a**) show chance levels (0.5) and in (**c**) show the maximum possible confidence value (3 = high confidence). Each dot represents one recording session. Black lines in (**a-c**) denote the mean results averaged across all recording sessions. One-way ANOVA between NB/SB/HB, degrees of freedom = (2, 57).

Extended Data Fig. 3. Boundary cells and event cells do not respond to clip onsets and clip offsets during encoding, Related to Fig. 3

a, Responses during the encoding stage from the same example boundary cells shown in Fig. 3a and Fig. 3b aligned to the clip onsets. **b,** Firing rates of all 42 boundary cells (solid and dashed arrows denote the examples in (**a**) during the encoding stage aligned to the clip onsets, averaged over trials within each boundary type and normalized to each neuron's maximum firing rate throughout the entire task (see color scale on bottom). **c,** Responses during the encoding stage from the same example boundary cells shown in (**a**) aligned to the clip offsets. **d**, Firing rates of all 42 boundary cells during the encoding stage aligned to the clip offsets using the same format as (**b**). **e,** Responses during the encoding stage from the same example event cells shown in Fig. 3e and Fig. 3f aligned to the clip onsets. **f,** Firing rates of all 36 event cells (solid and dashed arrows denote the examples in (**e**) during the encoding stage aligned to clip onsets, using the same format as (**b**). **g,** Responses during the encoding stage from the same example event cells shown in (**e**) aligned to the clip offsets. **h**, Firing rates of all 36 event cells during the encoding stage aligned to the clip offsets using the same format as (**b**). For (**a**), (**c**), (**e**), (**g**), Top: raster plot color coded for different boundary types (green: NB; blue: SB; red: HB). Bottom: Post-stimulus time histogram (bin size = 200ms, step size = 2ms, shaded areas represented ± s.e.m. across trials). (**b** and **f**) are copied from Fig. 3d and Fig. 3h for comparison purposes.
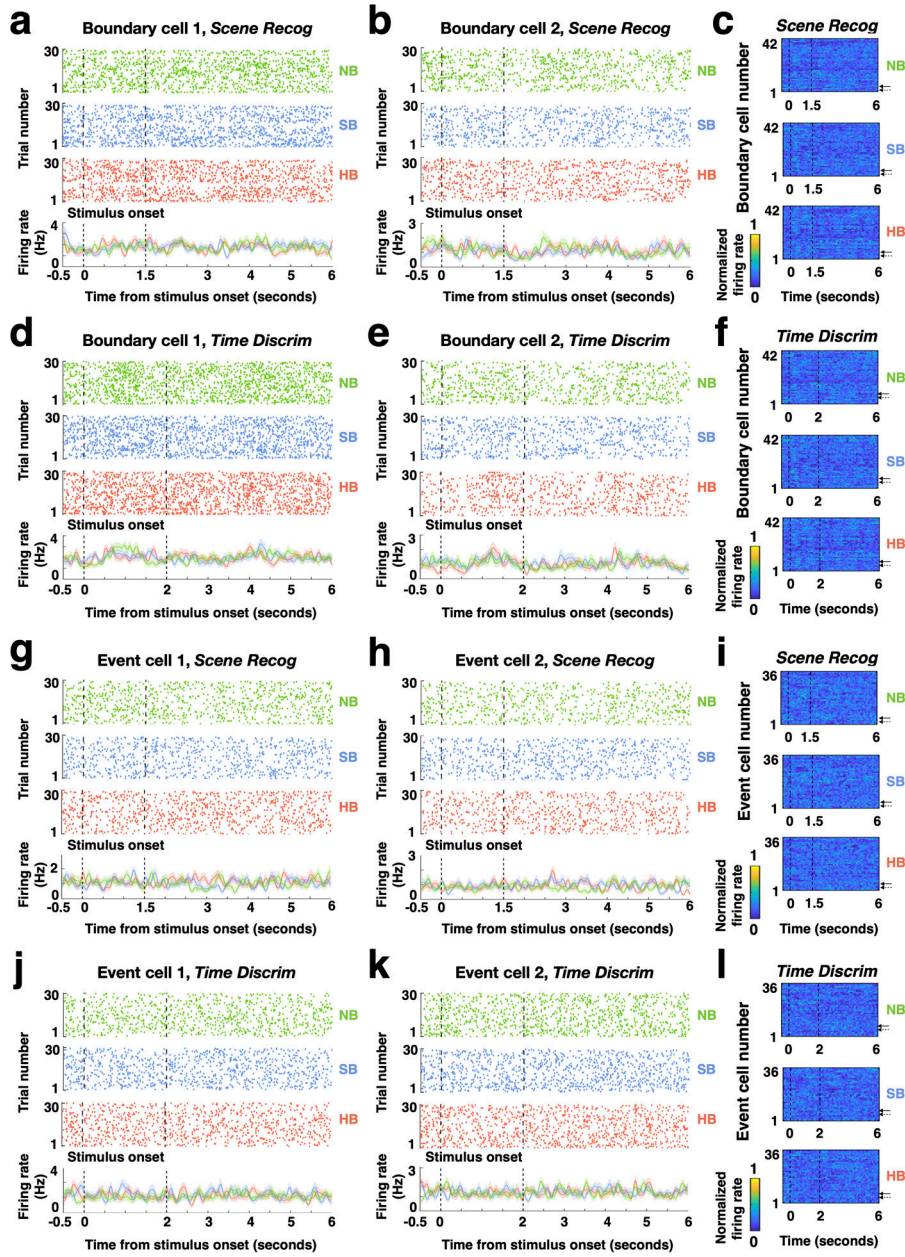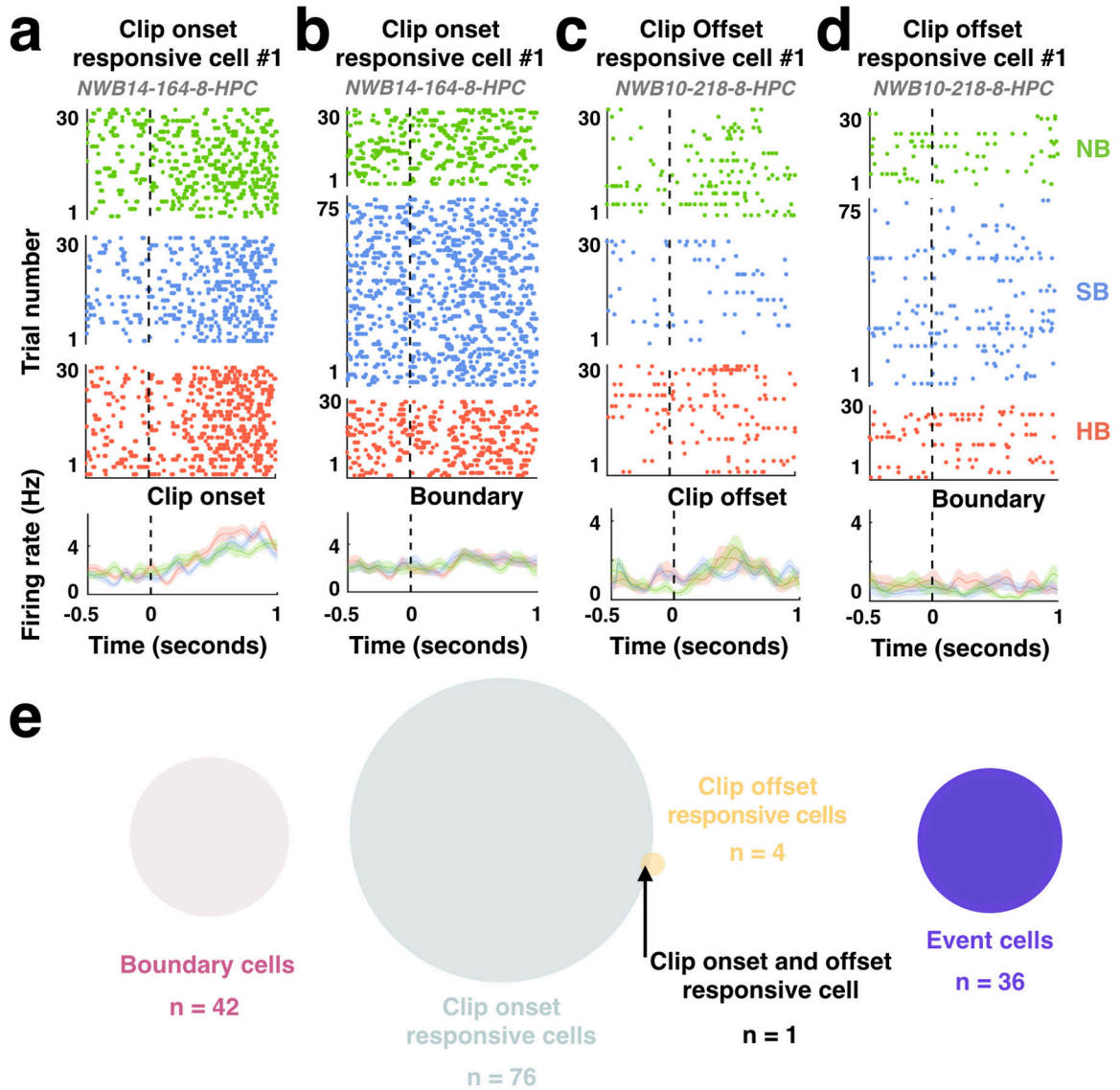
Extended Data Fig. 4. Boundary cells and event cells do not respond to image onsets and offsets during scene recognition and time discrimination, Related to Fig. 3

a-b, Responses during scene recognition from the same example boundary cells shown in Fig. 3a and Fig. 3b aligned to stimulus onset. **c,** Firing rates of all 42 boundary cells (solid and dashed arrows denote the examples in **a** and **b**) during scene recognition aligned to the stimulus onsets, averaged over trials within each boundary type and normalized to each neuron's maximum firing rate throughout the entire task (see color scale on bottom). **d-e**, Responses during time discrimination from the same example boundary cells shown in (**a** and **b**) aligned to stimulus onset. **f,** Firing rates of all 42 boundary cells during time discrimination aligned to the stimulus onset using the same format as in **c**. **g-h**, Responses during scene recognition from the same example event cells shown in Fig. 3e and Fig. 3f

aligned to stimulus onsets. **i,** Firing rates of all 36 event cells (solid and dashed arrows denote the examples in **g** and **h**) during scene recognition aligned to the stimulus onset, using the same format as in **a** and **b**. **j,** Responses during time discrimination from the same example event cells shown in **g** and **h** aligned to stimulus onset. **k,** Firing rates of all 36 event cells during time discrimination aligned to the stimulus onsets using the same format as in **f**. For (**a**), (**b**), (**d**), (**e**), (**g**), (**h**), (**j**), (**k**), Top: raster plot color coded for different boundary types (green: NB; blue: SB; red: HB). Bottom: Post-stimulus time histogram (bin size = 200ms, step size = 2ms, shaded areas represented ± s.e.m. across trials).


Extended Data Fig. 5. Neurons that respond to clip onsets and clip offsets do not overlap with boundary and event cells, Related to Fig. 3

a-b, Responses during the encoding stage from an example clip onset-responsive cell located in the amygdala aligned to clip onsets (**a**), and boundaries (**b**). Top: raster plots. Bottom: Post-stimulus time histogram (bin size = 200 ms, step size = 2ms, shaded areas represented

± s.e.m. across trials). A cell was considered as a clip onset cell if its firing rate differed significantly between a 1s window immediate before and after clip onset ($p < 0.05$, one-tailed permutation t-test). **c-d,** Responses during the encoding stage from an example clip offset-responsive cell located in the hippocampus aligned to clip offsets (**c**), and boundaries (**d**). A cell was considered as a clip offset cell if its firing rate differed significantly between a 1s window immediate before and after clip offsets ($p < 0.05$, one-tailed permutation t-test). Same format as (**a** and **b**). **e,** Seventy six out of 580 cells in the MTL qualified as clip onset-responsive cells and four out of 580 cells in the MTL qualified as clip offset-responsive cells. None of these were also selected as either boundary or event cells.
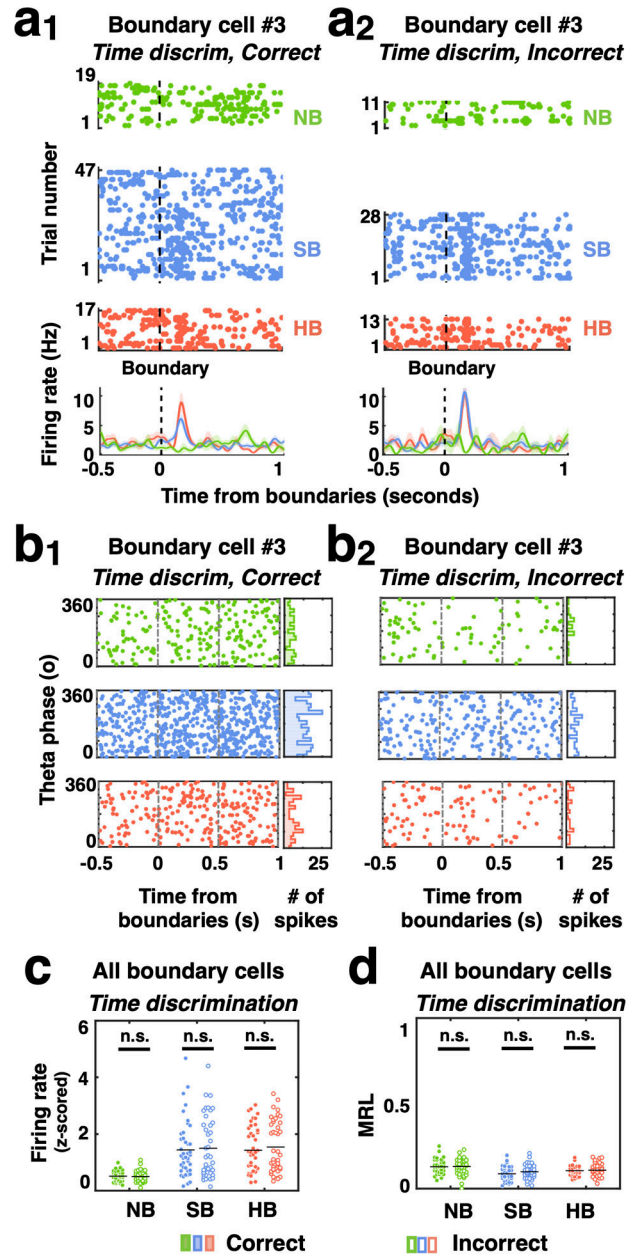
Extended Data Fig. 6. Responses of boundary cells during encoding grouped by memory outcomes from the time discrimination task, Related to Fig. 4

a_1-a_2, Response of the same example boundary cell in Fig. 4a and Fig. 4b. During encoding, this cell responded to SB and HB transitions regardless of whether the temporal order of the clip was later correctly (a_1) or incorrectly (a_2) retrieved in the time discrimination test. Shaded areas represented ± s.e.m. across trials. b_1- b_2, Left: timing of spikes from the same boundary cell shown in (a_1 and a_2) relative to theta phase calculated from the local field potentials, for clips whose temporal order were later correctly (b_1) or incorrectly (b_2) retrieved. Right: phase distribution of spike times within [0, 1] seconds time windows following the middle of the clip (NB) or boundary (SB, HB) for clips whose temporal order were later correctly (b_1) or incorrectly (b_2) retrieved. **c-d**, Population summary for all 42 boundary cells. **c,** Z-scored firing rate (0–1s after boundaries during encoding) for each boundary type did not differ between clips whose temporal orders were later correctly (color filled) vs. incorrectly (empty) retrieved. **d,** Mean resultant length (MRL) of spike times (relative to theta phases, 0–1s after boundaries during encoding) across all boundary cells for each boundary type did not differ between clips whose temporal orders were later correctly (color filled) vs. incorrectly (empty) retrieved. Each dot represents one boundary cell. Black lines in **c** and **d** denote the mean results averaged across all boundary cells. One-tailed permutation t-test, degrees of freedom = (1, 82).
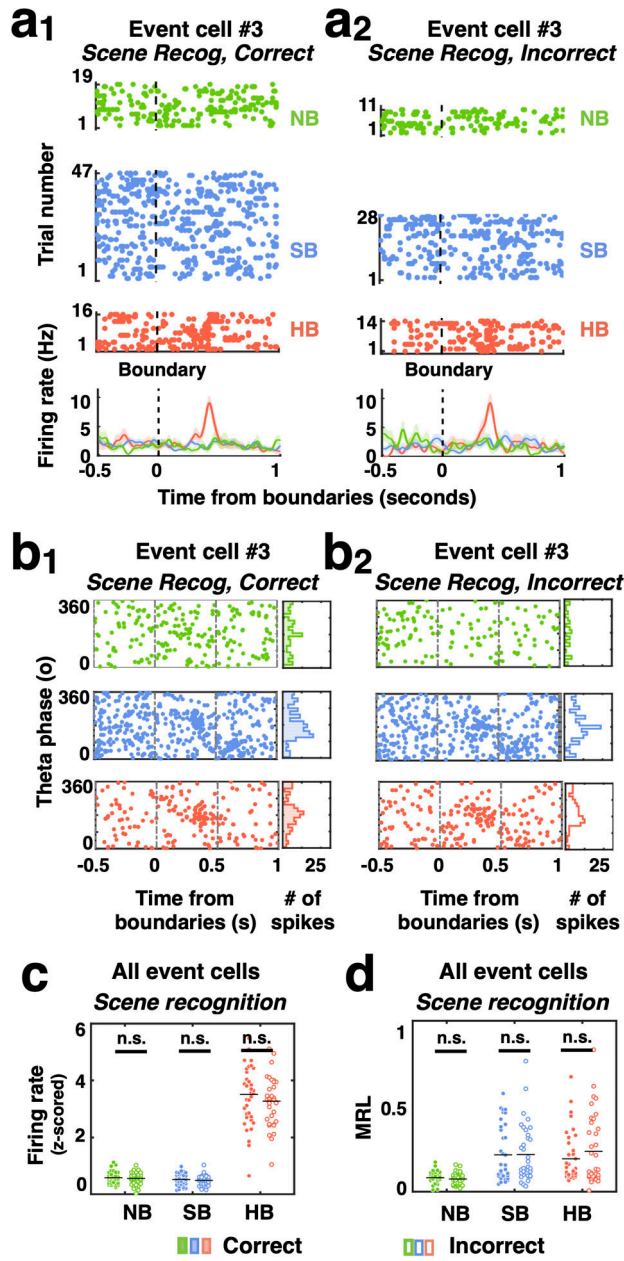
Extended Data Fig. 7. Responses of event cells during encoding grouped by memory outcomes from the scene recognition stage, Related to Fig. 4

a$_1$-a$_2$, Response of the same example event cell in Fig. 4e and Fig. 4f. During encoding, this cell responded to HB transitions regardless of whether frames were later correctly (**a$_1$**) or incorrectly (**a$_2$**) recognized in the scene recognition task. Shaded areas represented ± s.e.m. across trials. **b$_1$-b$_2$**, Left: timing of spikes from the same event cell shown in **a$_1$-a$_2$** relative to theta phase calculated from the local field potentials, for frames that were later correctly (**b$_1$**) or incorrectly (**b$_2$**) recognized. Right: phase distribution of spike times within [0, 1] seconds time windows following the middle of the clip (NB) or boundary (SB, HB) for frames that were later correctly (**b$_1$**) or incorrectly (**b$_2$**) recognized. **c-d**, Population summary for all 36 event cells. **c**, Z-scored firing rate (0–1s after boundaries during encoding) for

each boundary type did not differ between frames that were later correctly (color filled) vs. incorrectly (empty) recognized. **d**, Mean resultant length (MRL) of spike times (relative to theta phases, 0–1s after boundaries during encoding) across all event cells for each boundary type did not differ between frames that were later correctly (color filled) vs. incorrectly (empty) recognized. Each dot represents one event cell. Black lines in (**c**) and (**d**) denote the mean results averaged across all event cells (**c**, **d**). One-tailed permutation t-test, degree of freedom = (1, 70).
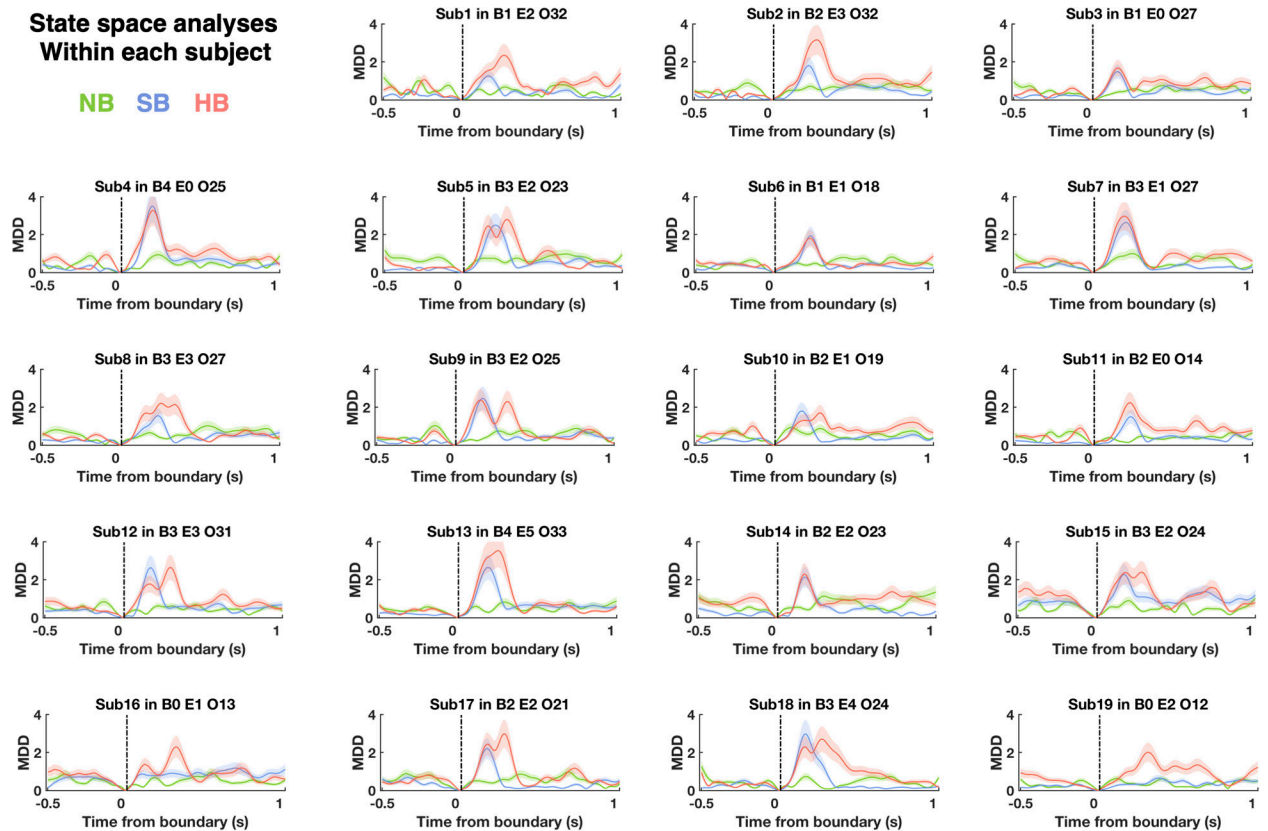
Extended Data Fig. 8. Neural state changes following soft and hard boundaries shown for individual subjects, Related to Fig. 5

Multidimensional distance (MDD, see Fig. 5d–g for definition) as a function of time aligned to the middle of the clip (green: NB) and boundaries (blue: SB, red: HB). MDD is shown for all MTL cells within each subject (e.g., "Sub1 in B1 E2 O32" denotes MDD computed by 1 boundary cell, 2 event cells and 32 other MTL cells in subject 1). Shaded areas represent ± s.e.m. across trials.
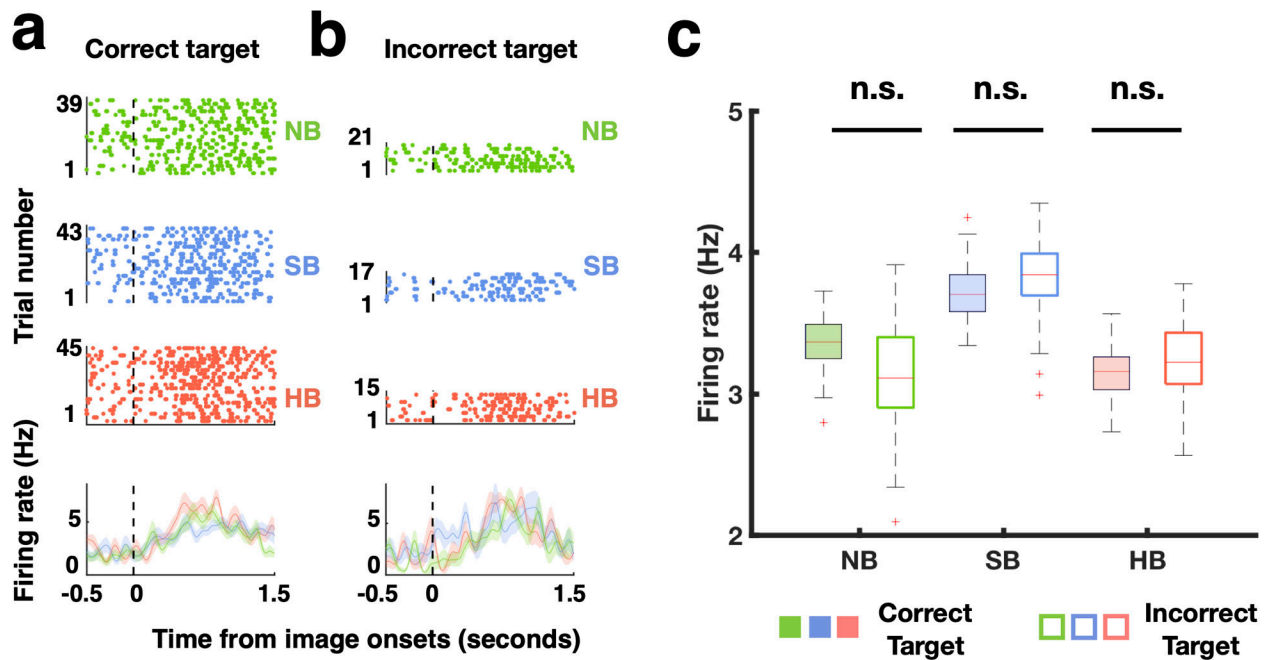
Extended Data Fig. 9. Clip-onsets responsive neurons respond to both correct and incorrect targets during scene recognition, Related to Fig. 6

a-b, Responses during scene recognition from an example clip onset-responsive cell (see definition in Extended Data Fig. 5) located in the amygdala aligned to image onsets in correctly recognized target (**a**) and forgotten target (**b**) trials. Top: raster plots. Bottom: Post-stimulus time histogram (bin size = 200 ms, step size = 2ms, shaded areas represented ± s.e.m. across trials). **c**, Comparison (across all 76 identified clip-onsets responsive neurons) between mean firing rates averaged within [0 1.5]s after image onsets for remembered vs forgotten targets. On each box, the central mark indicates the mean results averaged across all clip-onsets responsive neurons, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' marker symbol. One-way ANOVA, degrees of freedom = (1, 150)

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Data Availability

The data (in NWB format) that supports the key findings of this study are publicly available on the DANDI archive (https://doi.org/10.48324/dandi.000207/0.220216.0323)

References

1. Ezzyat Y & Davachi L What constitutes an episode in episodic memory? Psychol Sci 22, 243–252, doi:10.1177/0956797610393742 (2011). [PubMed: 21178116]

2. Tulving E Episodic memory: from mind to brain. Annu Rev Psychol 53, 1–25, doi:10.1146/annurev.psych.53.100901.135114 (2002). [PubMed: 11752477]

3. Radvansky GA & Zacks JM Event cognition. (Oxford University Press, 2014).

4. Desimone R & Duncan J Neural mechanisms of selective visual attention. Annu Rev Neurosci 18, 193–222, doi:10.1146/annurev.ne.18.030195.001205 (1995). [PubMed: 7605061]

5. Zacks JM et al. Human brain activity time-locked to perceptual event boundaries. Nat Neurosci 4, 651–655, doi:10.1038/88486 (2001). [PubMed: 11369948]

6. Avrahami J & Kareev Y The emergence of events. Cognition 53, 239–261, doi:10.1016/0010-0277(94)90050-7 (1994). [PubMed: 7842635]

7. DuBrow S & Davachi L The influence of context boundaries on memory for the sequential order of events. J Exp Psychol Gen 142, 1277–1286, doi:10.1037/a0034024 (2013). [PubMed: 23957281]

8. Chen J et al. Shared memories reveal shared structure in neural activity across individuals. Nat Neurosci 20, 115–125, doi:10.1038/nn.4450 (2017). [PubMed: 27918531]

9. Kurby CA & Zacks JM Segmentation in the perception and memory of events. Trends Cogn Sci 12, 72–79, doi:10.1016/j.tics.2007.11.004 (2008). [PubMed: 18178125]

10. Zacks JM, Speer NK, Swallow KM, Braver TS & Reynolds JR Event perception: a mind-brain perspective. Psychol Bull 133, 273–293, doi:10.1037/0033-2909.133.2.273 (2007). [PubMed: 17338600]

11. Jafarpour A, Griffin S, Lin JJ & Knight RT Medial Orbitofrontal Cortex, Dorsolateral Prefrontal Cortex, and Hippocampus Differentially Represent the Event Saliency. J Cogn Neurosci 31, 874–884, doi:10.1162/jocn_a_01392 (2019). [PubMed: 30883290]

12. Ben-Yakov A & Henson RN The Hippocampal Film Editor: Sensitivity and Specificity to Event Boundaries in Continuous Experience. J Neurosci 38, 10057–10068, doi:10.1523/JNEUROSCI.0524-18.2018 (2018). [PubMed: 30301758]

13. Baldassano C et al. Discovering Event Structure in Continuous Narrative Perception and Memory. Neuron 95, 709–721 e705, doi:10.1016/j.neuron.2017.06.041 (2017). [PubMed: 28772125]

14. Spiers HJ, Hayman RM, Jovalekic A, Marozzi E & Jeffery KJ Place field repetition and purely local remapping in a multicompartment environment. Cereb Cortex 25, 10–25, doi:10.1093/cercor/bht198 (2015). [PubMed: 23945240]

15. Derdikman D et al. Fragmentation of grid cell maps in a multicompartment environment. Nat Neurosci 12, 1325–1332, doi:10.1038/nn.2396 (2009). [PubMed: 19749749]

16. Lever C, Burton S, Jeewajee A, O'Keefe J & Burgess N Boundary vector cells in the subiculum of the hippocampal formation. J Neurosci 29, 9771–9777, doi:10.1523/JNEUROSCI.1319-09.2009 (2009). [PubMed: 19657030]

17. O'Keefe J & Burgess N Geometric determinants of the place fields of hippocampal neurons. Nature 381, 425–428, doi:10.1038/381425a0 (1996). [PubMed: 8632799]

18. Alme CB et al. Place cells in the hippocampus: eleven maps for eleven rooms. Proc Natl Acad Sci U S A 111, 18428–18435, doi:10.1073/pnas.1421056111 (2014). [PubMed: 25489089]

19. Colgin LL, Moser EI & Moser MB Understanding memory through hippocampal remapping. Trends Neurosci 31, 469–477, doi:10.1016/j.tins.2008.06.008 (2008). [PubMed: 18687478]

20. Grieves RM, Jenkins BW, Harland BC, Wood ER & Dudchenko PA Place field repetition and spatial learning in a multicompartment environment. Hippocampus 26, 118–134, doi:10.1002/hipo.22496 (2016). [PubMed: 26190393]

21. Sun C, Yang W, Martin J & Tonegawa S Hippocampal neurons represent events as transferable units of experience. Nat Neurosci 23, 651–663, doi:10.1038/s41593-020-0614-x (2020). [PubMed: 32251386]

22. Levine B et al. Episodic memory and the self in a case of isolated retrograde amnesia. Brain 121 (Pt 10), 1951–1973, doi:10.1093/brain/121.10.1951 (1998). [PubMed: 9798749]

23. Rutishauser U Testing Models of Human Declarative Memory at the Single-Neuron Level. Trends Cogn Sci 23, 510–524, doi:10.1016/j.tics.2019.03.006 (2019). [PubMed: 31031021]

24. Rutishauser U, Ross IB, Mamelak AN & Schuman EM Human memory strength is predicted by theta-frequency phase-locking of single neurons. Nature 464, 903–907, doi:10.1038/nature08860 (2010). [PubMed: 20336071]

25. Pacheco Estefan D et al. Coordinated representational reinstatement in the human hippocampus and lateral temporal cortex during episodic memory retrieval. Nat Commun 10, 2255, doi:10.1038/s41467-019-09569-0 (2019). [PubMed: 31113952]

26. Manning JR, Polyn SM, Baltuch GH, Litt B & Kahana MJ Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. Proc Natl Acad Sci U S A 108, 12893–12897, doi:10.1073/pnas.1015174108 (2011). [PubMed: 21737744]

27. Folkerts S, Rutishauser U & Howard MW Human Episodic Memory Retrieval Is Accompanied by a Neural Contiguity Effect. J Neurosci 38, 4200–4211, doi:10.1523/JNEUROSCI.2312-17.2018 (2018). [PubMed: 29615486]

28. Howard MW, Fotedar MS, Datey AV & Hasselmo ME The temporal context model in spatial navigation and relational learning: toward a common explanation of medial temporal lobe function across domains. Psychol Rev 112, 75–116, doi:10.1037/0033-295X.112.1.75 (2005). [PubMed: 15631589]

29. Swallow KM, Zacks JM & Abrams RA Event boundaries in perception affect memory encoding and updating. J Exp Psychol Gen 138, 236–257, doi:10.1037/a0015631 (2009). [PubMed: 19397382]

30. Richmond LL, Gold DA & Zacks JM Event perception: Translations and applications. J Appl Res Mem Cogn 6, 111–120, doi:10.1016/j.jarmac.2016.11.002 (2017). [PubMed: 28936393]

31. Ben-Yakov A & Dudai Y Constructing realistic engrams: poststimulus activity of hippocampus and dorsal striatum predicts subsequent episodic memory. J Neurosci 31, 9032–9042, doi:10.1523/JNEUROSCI.0702-11.2011 (2011). [PubMed: 21677186]

32. Ben-Yakov A, Eshel N & Dudai Y Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. J Exp Psychol Gen 142, 1255–1263, doi:10.1037/a0033558 (2013). [PubMed: 23815458]

33. Isik L, Singer J, Madsen JR, Kanwisher N & Kreiman G What is changing when: Decoding visual information in movies from human intracranial recordings. Neuroimage 180, 147–159, doi:10.1016/j.neuroimage.2017.08.027 (2018). [PubMed: 28823828]

34. Aminoff EM, Kveraga K & Bar M The role of the parahippocampal cortex in cognition. Trends Cogn Sci 17, 379–390, doi:10.1016/j.tics.2013.06.009 (2013). [PubMed: 23850264]

35. Lisman JE & Grace AA The hippocampal-VTA loop: controlling the entry of information into long-term memory. Neuron 46, 703–713 (2005). [PubMed: 15924857]

36. Vinogradova OS Hippocampus as comparator: role of the two input and two output systems of the hippocampus in selection and registration of information. Hippocampus 11, 578–598 (2001). [PubMed: 11732710]

37. Solstad T, Boccara CN, Kropff E, Moser MB & Moser EI Representation of geometric borders in the entorhinal cortex. Science 322, 1865–1868, doi:10.1126/science.1166466 (2008). [PubMed: 19095945]

38. Xiao X et al. Transformed Neural Pattern Reinstatement during Episodic Memory Retrieval. J Neurosci 37, 2986–2998, doi:10.1523/JNEUROSCI.2324-16.2017 (2017). [PubMed: 28202612]

39. Favila SE, Samide R, Sweigart SC & Kuhl BA Parietal Representations of Stimulus Features Are Amplified during Memory Retrieval and Flexibly Aligned with Top-Down Goals. J Neurosci 38, 7809–7821, doi:10.1523/JNEUROSCI.0564-18.2018 (2018). [PubMed: 30054390]

40. Jang AI, Wittig JH Jr., Inati SK & Zaghloul KA Human Cortical Neurons in the Anterior Temporal Lobe Reinstate Spiking Activity during Verbal Memory Retrieval. Curr Biol 27, 1700–1705 e1705, doi:10.1016/j.cub.2017.05.014 (2017). [PubMed: 28552361]

41. Howard MW & Natu VS Place from time: Reconstructing position from a distributed representation of temporal context. Neural Netw 18, 1150–1162, doi:10.1016/j.neunet.2005.08.002 (2005). [PubMed: 16198538]

42. Polyn SM & Kahana MJ Memory search and the neural representation of context. Trends Cogn Sci 12, 24–30, doi:10.1016/j.tics.2007.10.010 (2008). [PubMed: 18069046]

43. Krizhevsky A, Sutskever I & Hinton GE Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 1097–1105 (2012).

44. Rutishauser U, Schuman EM & Mamelak AN Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, in vivo. J Neurosci Methods 154, 204–224, doi:10.1016/j.jneumeth.2005.12.033 (2006). [PubMed: 16488479]

45. Fried I, Rutishauser U, Cerf M & Kreiman G Single neuron studies of the human brain : probing cognition. (The MIT Press, 2014).

46. Kaminski J et al. Persistently active neurons in human medial frontal and medial temporal lobe support working memory. Nat Neurosci 20, 590–601, doi:10.1038/nn.4509 (2017). [PubMed: 28218914]

47. Pouzat C, Mazor O & Laurent G Using noise signature to optimize spike-sorting and to assess neuronal classification quality. J Neurosci Methods 122, 43–57, doi:10.1016/s0165-0270(02)00276-5 (2002). [PubMed: 12535763]

48. Harris KD, Henze DA, Csicsvari J, Hirase H & Buzsaki G Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. J Neurophysiol 84, 401–414, doi:10.1152/jn.2000.84.1.401 (2000). [PubMed: 10899214]

49. Reuter M, Rosas HD & Fischl B Highly accurate inverse consistent registration: a robust approach. Neuroimage 53, 1181–1196, doi:10.1016/j.neuroimage.2010.07.020 (2010). [PubMed: 20637289]

50. Pauli WM, Nili AN & Tyszka JM A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. Sci Data 5, 180063, doi:10.1038/sdata.2018.63 (2018). [PubMed: 29664465]

51. Avants B et al. Multivariate analysis of structural and diffusion imaging in traumatic brain injury. Acad Radiol 15, 1360–1375, doi:10.1016/j.acra.2008.07.007 (2008). [PubMed: 18995188]

52. Banaie Boroujeni K, Tiesinga P & Womelsdorf T Adaptive spike-artifact removal from local field potentials uncovers prominent beta and gamma band neuronal synchronization. J Neurosci Methods 330, 108485, doi:10.1016/j.jneumeth.2019.108485 (2020). [PubMed: 31705936]

53. Oostenveld R, Fries P, Maris E & Schoffelen JM FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011, 156869, doi:10.1155/2011/156869 (2011). [PubMed: 21253357]

54. Delorme A & Makeig S EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134, 9–21 (2004). [PubMed: 15102499]

55. Kobak D et al. Demixed principal component analysis of neural population data. Elife 5, doi:10.7554/eLife.10989 (2016).

56. Bausch M et al. Concept neurons in the human medial temporal lobe flexibly represent abstract relations between concepts. Nat Commun 12, 6164, doi:10.1038/S41467-021-26327-3 (2021). [PubMed: 34697305]
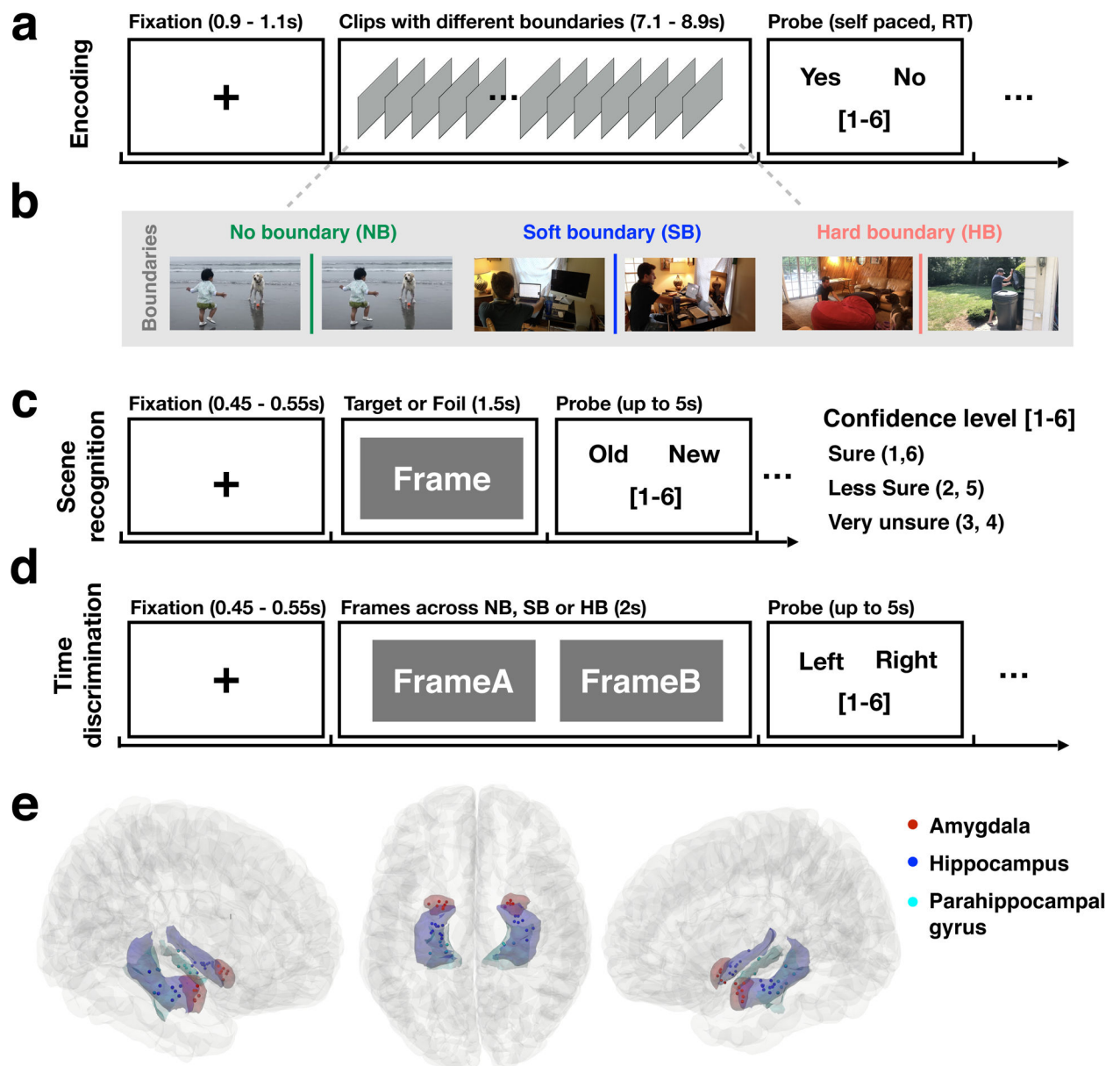
Fig. 1: Experiment and recording locations.
a, Encoding task. Subjects watched 90 video clips (~ 8 seconds each, no audio) with either no boundary (NB, continuous movie shot), a soft boundary (SB, cut to a new scene within the same movie, 1 to 3 SB per clip), or a hard boundary (HB, cut to a different movie, 1 HB per clip). Every 4–8 clips, subjects were prompted to answer a Yes/No question related to the clip content, together with a confidence rating (Methods). RT = reaction time **b,** Example boundaries (visual features of boundaries in Supplementary Table 1). Owing to copyright restrictions, the images shown are different from those used for the experiment. **c,** Scene recognition memory task. Subjects indicated whether a static image was New or Old (seen during encoding task), together with a confidence rating. **d,** Time discrimination task. Subjects indicated which of two frames they saw first during the encoding task, together with a confidence rating. **e,** Recording locations of the 39 microwire bundles that contained
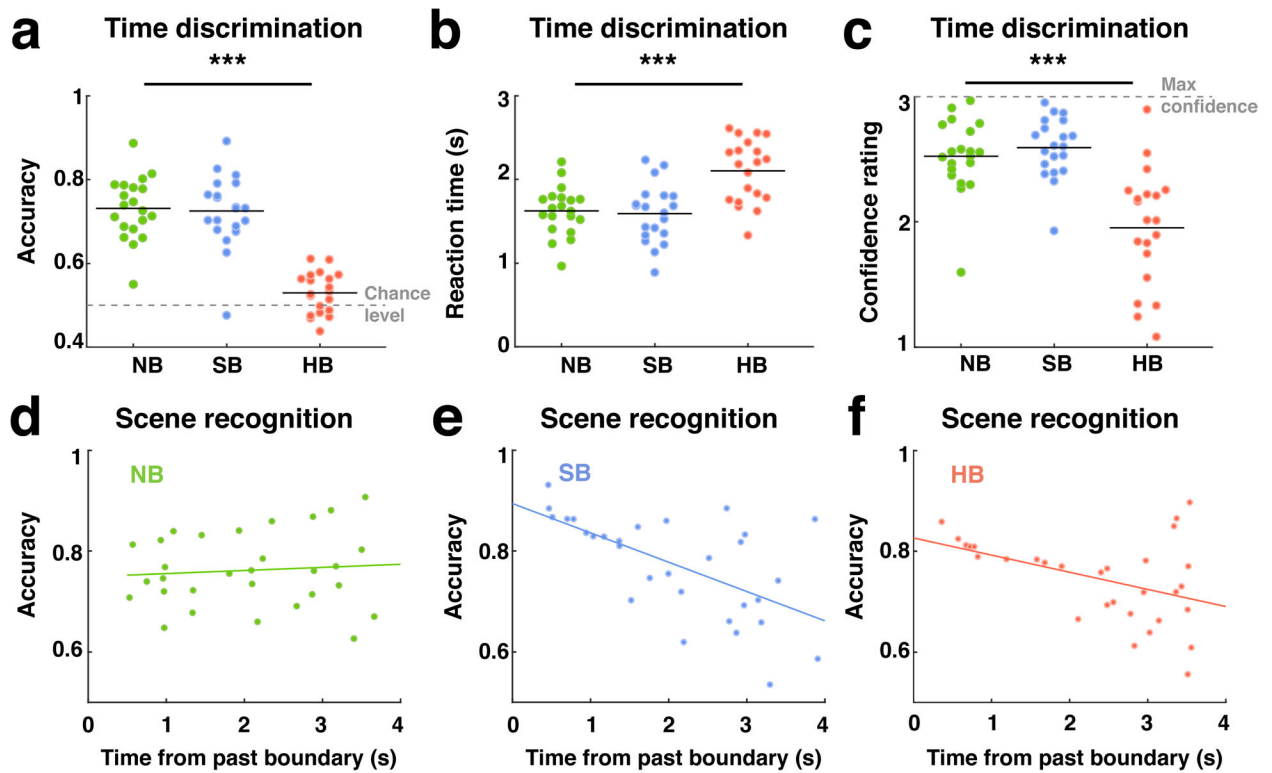
at least one boundary/event neuron (see MNI coordinates in Supplementary Table 3 and Extended Data Fig. 1) across all subjects (subject information in Supplementary Table 2) in the amygdala (red), hippocampus (blue), or parahippocampal gyrus (cyan), rendered on a template brain. Each dot represents the location of a microwire bundle.

Fig. 2: Behavior.

Hard boundaries impaired time discrimination memory while soft and hard boundaries improved scene recognition memory for frames close to them. **a-c**, Performance in time discrimination task (see also Supplementary Figs. 1–2) quantified by accuracy in **a** (F (2, 57) = 51.33, p = 2×10^{-13}; one-tailed ANOVA), reaction time in **b** (F(2, 57) = 14.25 p = 10×10^{-6}; one-tailed ANOVA), and mean confidence level in **c** (F(2, 57) = 20.41, p = 2×10^{-7}; one-tailed ANOVA) across all the trials for NB (green), SB (blue), and HB (red) trials. Behavior data for the scene recognition task is shown in Extended Data Fig. 2. **d-f,** Scene recognition accuracy as a function of time elapsed between the target frame and its nearest past boundary (distance effect for time discrimination accuracy and future boundaries shown in Supplementary Fig. 3) plotted separately for NB (**d**), SB (**e**) and HB (**f**). For NB clips, time from the past boundary is measured relative to the middle of the clip. Each dot represents one recording session in (**a-c**) and one clip in (**d-f**). Black lines in (**a-c**) denote the mean of the results and colored lines in (**d-f**) are the fitted lines for linear regression. ***$P < 0.001$
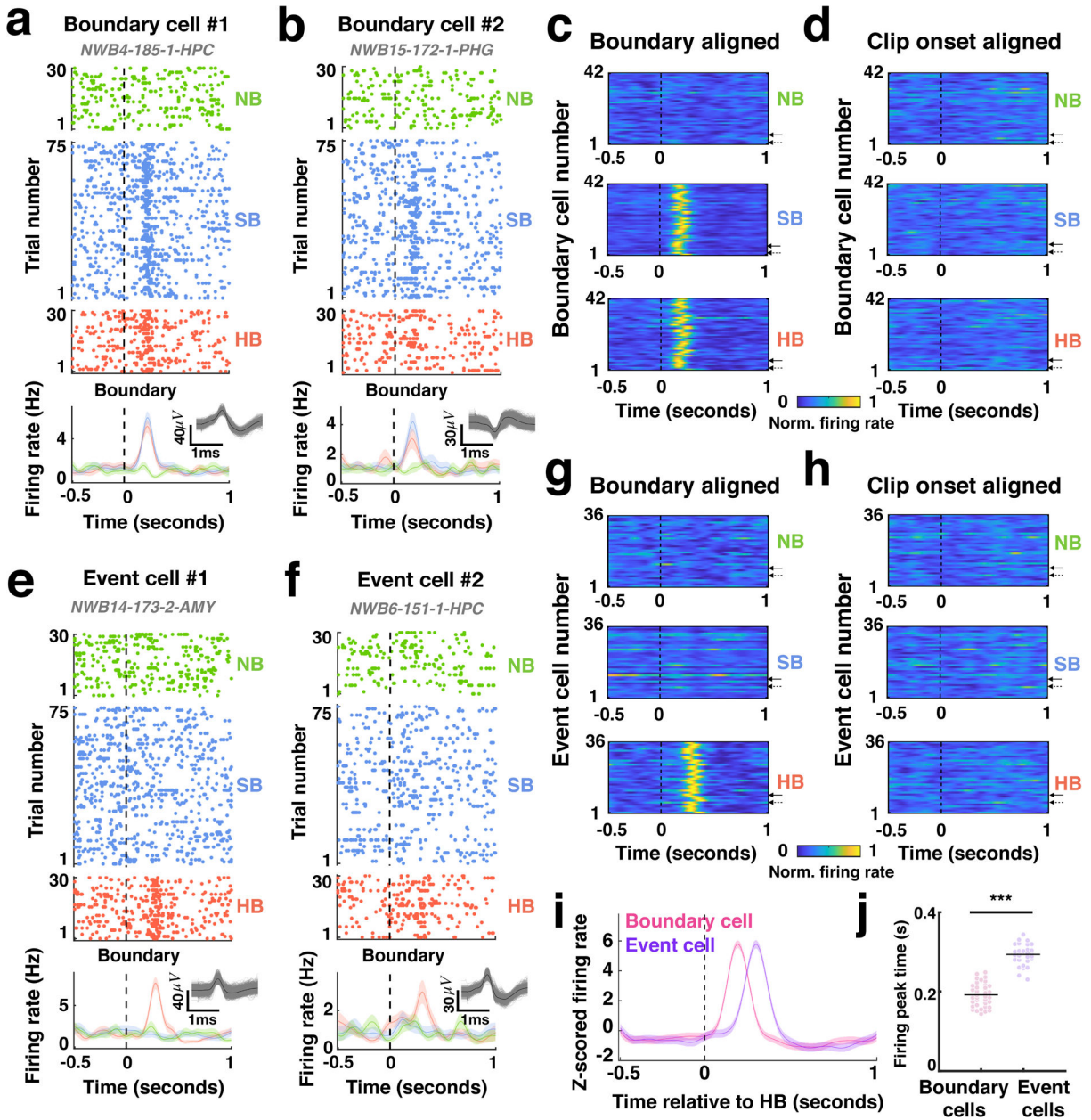
Fig. 3: Boundary cells and event cells demarcate different types of episodic transitions.
a-b, Responses during the encoding stage from two example boundary cells located in the parahippocampal gyrus and hippocampus, respectively (spike sorting quality of all detected cells shown in Supplementary Fig. 5). Boundary cells responded to both SB (blue) and HB (red) transitions. Responses are aligned to the middle point of the clip (NB, green) or to the boundary (SB, HB). Top: raster plots. Bottom: Post-stimulus time histogram (bin size = 200 ms, step size = 2ms, shaded areas represented ± s.e.m. across trials). Insets: all spike extracellular waveforms (gray) and mean (black). **c-d**, Firing rates of all 42 boundary cells (solid and dashed arrows denote the examples in **a** and **b**, respectively) during the encoding stage aligned to the boundaries in **c** or clip onsets in **d**, averaged over trials within each boundary type and normalized to each neuron's maximum firing rate from the entire task

recording (see color scale on bottom). **e-f**, Responses during the encoding stage from two example event cells located in the hippocampus and amygdala, respectively. Event cells respond to HB (red) but not SB (blue) nor NB transitions. Post-stimulus time histogram (bin size = 200 ms, step size = 2ms, shaded areas represented ± s.e.m. across trials). **g-h**, Firing rates of all 36 event cells (solid and dashed arrows denote the examples in **e** and **f**, respectively) during the encoding stage, using the same format as in **c** and **d**. Both boundary cells and event cells in the medial temporal lobe do not respond to the clip onsets (**d**, **h**) and clip offsets (Extended Data Fig. 3) during encoding, and image onsets and offsets during scene recognition and time discrimination (Extended Data Fig. 4). No significant difference in saccades was found after clip onsets vs after boundary transitions for one subject where we could record eye movement data simultaneously with the neurophysiological data (Supplementary Fig. 7). **i,** Latency analysis. Firing rate during HB transitions (to which both boundary cells and event cells responded) reached peak response earlier for boundary cells (pink) compared to event cells (purple). Shown is average z-scored firing rate normalized using the average and standard deviation of the firing rates and aligned to HB (bin size = 200 ms, step size = 2ms, shaded areas represented ± s.e.m. across all boundary cells or event cells). **j,** Peak times of average firing rate traces of all boundary cells (pink) and all event cells (purple) ($F(1, 76) = 274.78$, $p = 6 \times 10^{-27}$, one-tailed ANOVA). Each dot represents one boundary cell (pink) or one event cell (purple). Black lines denote the mean averaged across all boundary cells or event cells. ***$P < 0.001$, one-way ANOVA, degrees of freedom = (1, 76). The spatial distribution of boundary cells and event cells is shown in Supplementary Table 4.
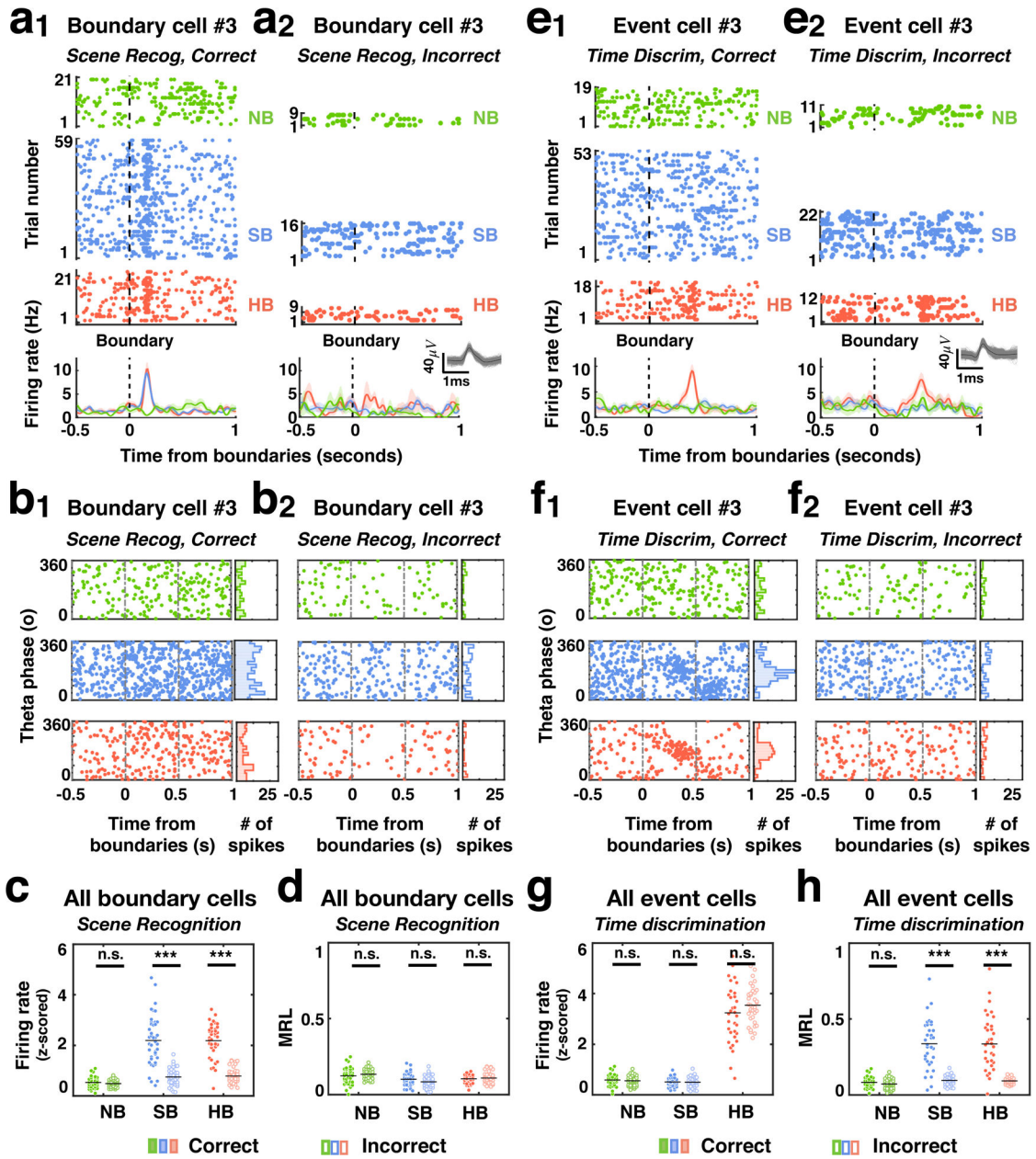
Fig. 4: Responses of boundary cells and event cells during encoding correlate with later retrieval success.

a-d, Response of boundary cells during encoding grouped by subjects' subsequent memory performance in the scene recognition task. **a₁-a₂**, Boundary cell recorded in the hippocampus. During encoding, this cell responded more strongly to SB and HB transitions than NB if the frame following the boundary in that trial was correctly identified during the scene recognition task (**a₁**) compared to incorrect trials (**a₂**). Format as in Fig. 3. Shaded areas represented ± s.e.m. across trials. **b₁-b₂**, Left: timing of spikes from the same boundary cell shown in (**a**) relative to theta phase calculated from the local field potentials, for clips of which frames were later remembered (**b₁**) or forgotten (**b₂**). Right: phase distribution of spike times in the 1s period following the middle of the clip (NB)

or boundary (SB, HB) for clips from which frames were remembered ($\mathbf{b_1}$) and forgotten ($\mathbf{b_2}$). **c-d,** Population summary for all 42 boundary cells. Black lines denote the mean results averaged across all 42 boundary cells. **c**, Z-scored firing rate (0–1s after boundaries during encoding) differed significantly between boundaries after which frames were remembered (color filled) vs. forgotten (empty) for both SB and HB (SB: $F(1, 82) = 82.93$, $p = 4 \times 10^{-14}$; HB: $F(1, 82) = 156.9$, $p = 1 \times 10^{-20}$; NB: $F(1, 82) = 1.18$, $p = 0.28$; one-tailed ANOVA). **d**, Mean resultant length (MRL) of spike times (i.e., sum of vectors with vector lengths equal to 1 and vector angles equal to the spike timings relative to theta phases 0–1s after boundaries during encoding, divided by total number of vectors; value range [0 1]: 0 = uniform distribution, i.e., neurons fire at random theta phases; 1 = unimodal distribution, i.e., neurons fire at the same theta phase) across all boundary cells for each boundary type did not differ significantly between correct (color filled) and incorrect (empty) clips. **e-h**, Response of boundary cells during encoding grouped by subjects' subsequent memory performance in the time discrimination task. $\mathbf{e_1}$-$\mathbf{e_2}$, Example event cell recorded in the hippocampus that responded to HB transition regardless of whether the temporal order of the clip was later correctly ($\mathbf{e_1}$) or incorrectly ($\mathbf{e_2}$) recalled in the time discrimination task. Shaded areas represented ± s.e.m. across trials. Format same as in **a** but clips were grouped based on memory outcomes in the time discrimination task. $\mathbf{f_1}$- $\mathbf{f_2}$, The spike timing of the same event cell shown in ($\mathbf{e_1}$-$\mathbf{e_2}$) relative to theta phase plotted for correct ($\mathbf{f_1}$) and incorrect ($\mathbf{f_2}$) trials. Format same as in **b** but clips were grouped based on memory outcomes in the time discrimination task. **g-h**, Population summary for all 36 event cells. Black lines denote the mean results averaged across all 36 event cells. **g**, Z-scored firing rate (0–1s after boundaries during encoding) did not differ significantly between later correctly (color filled) or incorrectly (empty) remembered temporal orders for all three boundary types. **h,** MRL of spike times (relative to theta phases, 0–1s after boundaries during encoding) was significantly larger after SB and HB transitions if the temporal order of the clip was correctly recalled (color filled) compared to the incorrect ones (empty) (SB: $F(1, 70) = 81.55$, $p = 2 \times 10^{-13}$; HB: $F(1, 70) = 60.79$, $p = 4 \times 10^{-11}$; NB: $F(1, 70) = 1.53$, $p = 0.22$; one-tailed ANOVA). Each dot represents one boundary cell (in **c** and **d**) or one event cell (in **g** and **h**). Black lines (in **c, d, g, h**) denote the mean of the results. Note that in (**a-d**) the neural responses of boundary cells reflect whether subjects remembered or forgot target frames that *followed* a boundary. Results computed based on trials grouped by subjects' memory performance for target frame *before* a boundary are shown in Supplementary Fig. 8. ***$P < 0.001$, one-way ANOVA, degrees of freedom = (1, 82) for (**c** and **d**) and degrees of freedom = (1, 70) for (**g** and **h**).
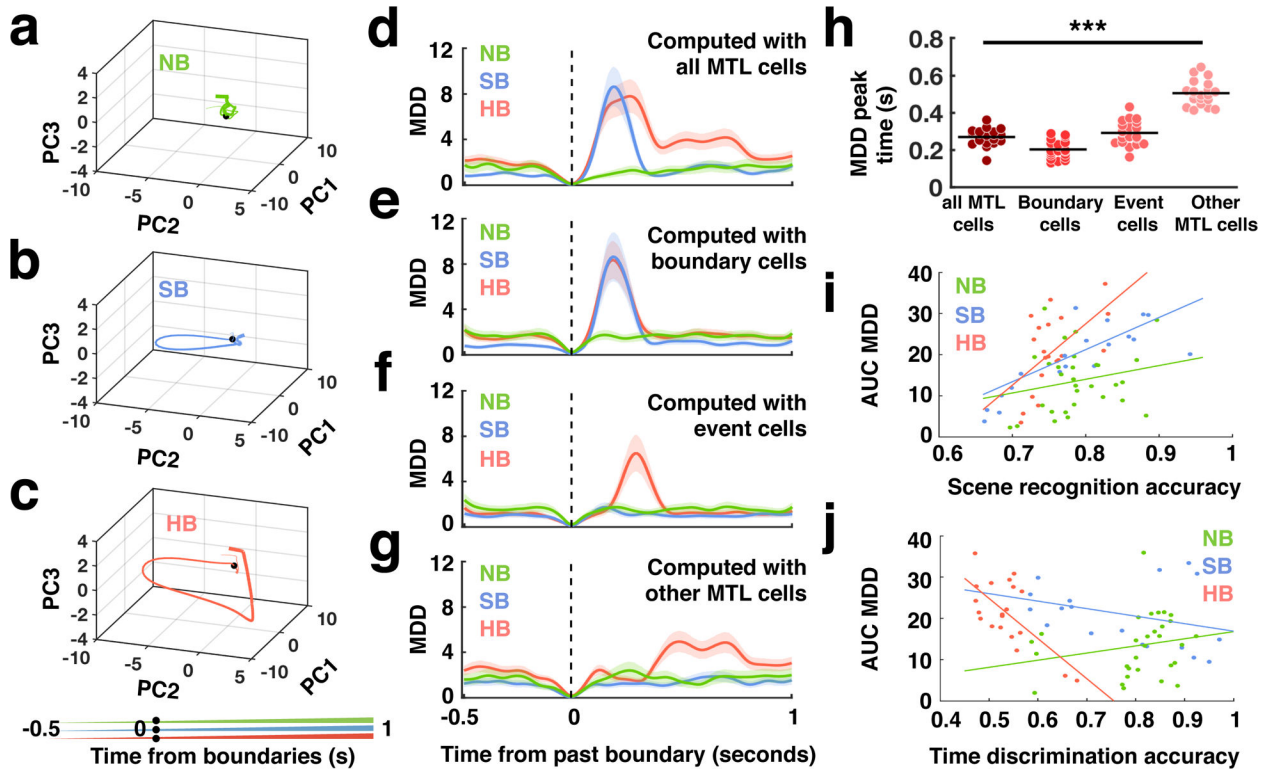
Fig. 5: Population neural state shift magnitude following episodic transitions reflects subjects' subsequent memory performance.

a-c, Trajectories in neural state space formed by the top three principal components (PCs with most explained variance: PC1 = 26.05%, PC2 = 10.89%; PC3 = 6.69%) summarizing the activity of all MTL cells during the encoding stage for clips containing NB (**a**), SB (**b**) and HB (**b**). Each data point indicates the neural state at a specific time relative to boundary onset (line thickness indicates time; see scale on bottom). Black dots mark the time of the boundary (SB, HB) or the middle of the clip (NB). **d-g,** Multidimensional distance (MDD, i.e., Euclidean distance relative to boundaries in the PC space formed by all PCs that cover explained variance 99%) as a function of time aligned to the middle of the clip (green: NB) and boundaries (blue: SB, red: HB). MDD is shown for all MTL cells (**d**; n = 580 in top 55 PCs space), all boundary cells (**e**; n = 42 in top 27 PCs space), all event cells (**f**; n = 36 in top 26 PCs space), and all other MTL cells (i.e., non-boundary/event cells in the MTL; **g**; n = 502 in top 58 PCs space). Shaded areas represent ± s.e.m. across trials. See neural state shifts within each subject in Extended Data Fig. 8. **h,** Latency analysis. Time when MDD shown in (**d-g**) reached peak value following HB (red lines) significantly differed when computed with different groups of cells ($F(3, 76) = 103.96$, $p = 8 \times 10^{-27}$, one-tailed ANOVA). Black lines denote the mean results averaged across different cell populations. **i-j,** Correlation between distance traveled in state space following boundaries and behavior. **i,** Positive correlation between AUC MDD (sum of Euclidean distances within [0, 1] seconds time window after boundaries in the PC space) and scene recognition accuracy. Dots mark the accuracy in the scene recognition task (x-axis) and the AUC MDD during encoding (y-axis) of the target frames plotted separately for frames that follow NB (green: $r = 0.214$, $p = 0.256$, Pearson correlation), SB (blue: $r = 0.653$, $p = 0.002$, Pearson correlation) and

HB (red: $r = 0.565$, $p = 0.009$, Pearson correlation). **j,** Negative Correlation between the AUC MDD versus time discrimination accuracy plotted in the same format as (**i**) for NB (green: $r = 0.212$, $p = 0.261$, Pearson correlation), SB (blue: $r = -0.273$, $p = 0.244$, Pearson correlation) and HB (red: $r = -0.677$, $p = 0.001$, Pearson correlation).***$P < 0.001$, one-way ANOVA, degrees of freedom = (3, 72) in (**h**).
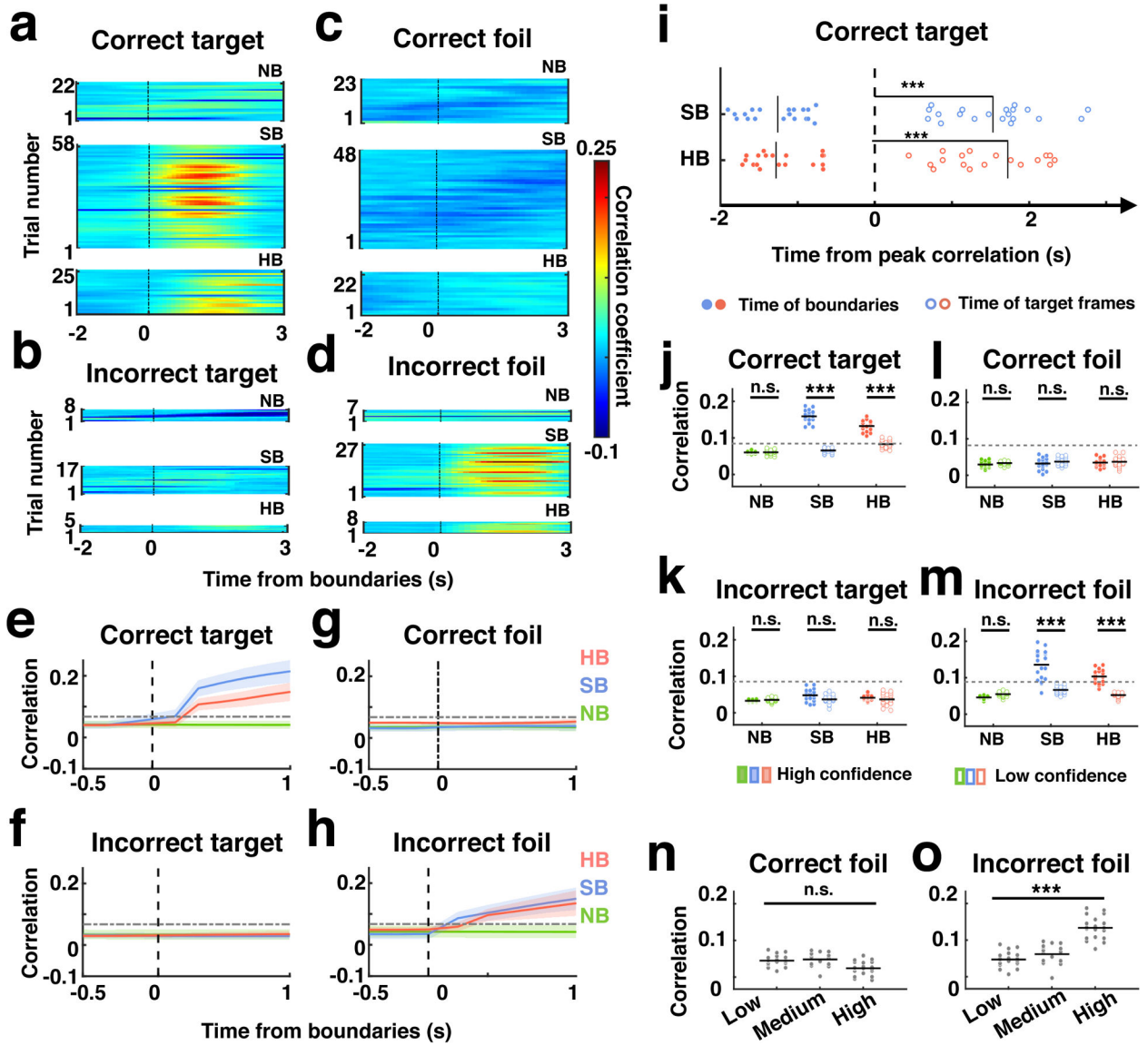
Fig. 6: Reinstatement of neural context after boundaries during recognition.
a-d, Single-subject example. Color code indicates correlation between the population response during scene recognition (0–1.5s relative to stimulus onset) and the encoding period (sliding window of 1.5s and 100ms step size). Correlations are aligned to the middle of the clip (NB) or boundaries (SB, HB) and are shown separately for correctly recognized familiar target (**a**), correctly recognized novel (not seen) foils (**c**), forgotten target (**b**) and incorrectly recognized foils (false positives, **d**) in the scene recognition task. See the correlation plots for the rest of the subjects in Supplementary Figs. 10–11. **e-h,** Population summary. Correlation coefficient as shown in part (**a-d**), averaged across all subjects for NB (green), SB (blue), and HB (red) trials. Shaded areas represented ± s.e.m. across subjects. The grey dashed horizontal lines denote the significant threshold (p < 0.01, one-tailed permutation test, Methods). See the same analyses after excluding boundary cells and event cells and only for boundary cells and event cells (Supplementary Fig. 12). **i,** The reinstated

neural context was located in between the boundary and the tested frame. For trials with target frames extracted after boundaries, the time distance from when the correlation coefficient peaks to the time of SB and HB (filled circles: SB: -1.26 ± 0.38 seconds, $t_{19} = 14.68$, $p = 8\times10^{-12}$; HB: -1.28 ± 0.48 seconds, $t_{19} = 11.80$, $p = 3\times10^{-10}$; one-tailed t-test) or target frames (empty circles; SB: 1.53 ± 0.61 seconds, $t_{19} = 11.18$, $p = 8\times10^{-10}$; HB: 1.72 ± 1.03 seconds, $t_{19} = 7.44$, $p = 5\times10^{-7}$; one-tailed t-test). Negative/Positive values denote the point of time of boundaries (negative) or target frames (positive) relative to when the correlation coefficient reaches its peak value. Asterisks indicate the significance of the peak correlation leading the time of target frames. See the same analyses with correlation computed using different window sizes (Supplementary Fig. 13). **j-m**, Population summary (confidence). Reinstatement differed between frames remembered with high (filled circles) and low (empty circles) confidence responses for "old" decisions (correct targets and incorrect foils) in SB and HB conditions but not 'new' decisions (correct foils and incorrect targets) and NB conditions, regardless of whether they were correct or incorrect (SB, correct targets: $p = 5\times10^{-10}$; HB, correct targets: $p = 4\times10^{-6}$; NB, correct targets: $p = 0.79$; SB, incorrect foils: $p = 5\times10^{-7}$; HB, incorrect foils: $p = 5\times10^{-5}$; NB, incorrect foils: $p = 0.18$; one-tailed t-test). Correlation coefficients as shown in part (**e-h**), averaged over $[0–1]$s after boundaries. **n-o**, Population summary (target-foil similarity). Correlation coefficients versus similarity ratings between targets and foils, plotted for correct (**n**; $F(2, 54) = 2.182$, $p = 0.144$; one-tailed ANOVA) and incorrect recognized foils (**o**; $F(2, 54) = 10.67$, $p = 1\times10^{-4}$; one-tailed ANOVA). Each dot represents one recording session. Black lines in (**i-o**) denote the mean results averaged across all recording sessions. $***P < 0.001$.