Title
Emotional Memory Processing and Pattern Separation in Medial Temporal Lobe

Author
Zhang, Haoxin

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Emotional Memory Processing and Pattern Separation in Medial Temporal Lobe

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Engineering

by

Haoxin Zhang

Dissertation Committee:
Professor Jack J. Lin, Chair
Professor Robert T. Knight
Professor Beth A. Lapour

2022

DEDICATION

To

My Mom, Liu
My Dad, Zhang
My future loved one

An experience may be so exciting emotionally as almost to leave a scar on the cerebral tissues.
William James (1842-1910)
The Principles of Psychology (1890)

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to thank my advisor, Jack Lin for his outstanding mentorship, fearless enthusiasm, and persistent support. I thank Robert Knight for his insightful feedback and guidance.

I thank Ivan Skelin for his knowledgeable and thoughtful advice. I thank Jie Zheng, Shiting Ma, Rebecca Stevenson and Logan Harriger for their assistance with data collection, electrode localization and sharing methodological brilliance with me.

I thank Michelle Paff, Sumeet Vadera and Lilit Mnatsakanyan for patient care, the staff at the Epilepsy Center at University of California Irvine Medical Center, and the patients who participated in the research.

This work was supported by Bridge Fellowship, University of California Irvine, U19NS107609-01 (PI: Knight, subcontract to R.T.K), U01-NS108916 (PI: J.J.L).

VITA

Haoxin Zhang, Ph.D. Candidate.

EDUCATION

2016-Current	University of California Irvine, Irvine, CA Doctorate in Philosophy Biomedical Engineering Mentor: Jack J. Lin, MD
2012-2016	University of Shanghai for Science and Technology, Shanghai, China Bachelor of Science Biomedical Engineering

ACADEMIC APPOINTMENT

2017 – Current	Graduate Student Researcher, Department of Biomedical Engineering, Henry Samueli School of Engineering, UC Irvine

LEADERSHIP

2020 – Current	Founder and President, Spike neuroscience journal club, UC Irvine

HONORS AND ACHIEVEMENTS

2017	Engineering Bridge Fellowship, University of California Irvine
2016	Excellent Undergraduate Thesis Award, University of Shanghai for Science and Technology, Shanghai, China
2014 and 2015	Excellent Study Award, University of Shanghai for Science and Technology, Shanghai, China
2012	Second Prize, 27th Shanghai Adolescents Science & Technology Innovation Fair, Shanghai, China
2012	Second Prize, Hope Cup National Mathematics Invitational Tournament, Shanghai, China

PUBLICATIONS

1. Skelin Ivan, **Haoxin Zhang**, Jie Zheng, Shiting Ma, Bryce A. Mander, Olivia Kim Mcmanus, Sumeet Vadera, Robert T. Knight, Bruce L. McNaughton, and Jack J. Lin. "Coupling between slow waves and sharp-wave ripples engages distributed

neural activity during sleep in humans." *Proceedings of the National Academy of Sciences* 118, no. 21 (2021).

2. Berken Utku Demirel, Ivan Skelin, **Haoxin Zhang**, Jack J. Lin, Mohammad Abdullah Al Faruque. "Single-Channel EEG Based Arousal Level Estimation Using Multitaper Spectrum Estimation at Low-Power Wearable Devices." *43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society* (in Press).
3. **Haoxin Zhang**, Ivan Skelin, Shiting Ma, Michelle Paff, Michael A. Yassa, Robert T. Knight and Jack J. Lin. "Awake ripple enhance emotional memory encoding in the human brain." (under review, Nature Neuroscience, 2021)
4. **Haoxin Zhang**, Ivan Skelin, Shiting Ma, Jack J. Lin. "Subsecond dynamics of behaviorally-relevant pattern separation in the human hippocampus." (in preparation)

ABSTRACTS

1. **Haoxin Zhang**, Ivan Skelin, Shiting Ma, Michelle Paff, Michael A. Yassa, Robert T. Knight, Jack J. Lin. Awake ripples enhance emotional memory encoding in the human brain. Society for Neuroscience. Washington, DC, 2021
2. Ivan Skelin*, **Haoxin Zhang***, Shiting Ma, Sumeet Vadera, Michelle Paff, Indranil Sen-Gupta, Lilit Mnatsakanyan, Frank P. K. Hsu, Elizabeth A. Buffalo, Robert T. Knight,Jack J. Lin. Phase dependent action outcome and rule representations observed in single neurons in human brain. BRAIN Initiative Meeting, 2020
3. **Haoxin Zhang**, Ivan Skelin, Shiting Ma, Michael A. Yassa, Jack J. Lin. Dimensionality reduction in human hippocampus (dentate/CA3, CA1) support pattern separation. Society for Neuroscience. Chicago IL, 2019
4. Jack J. Lin, **Haoxin Zhang**, Bruce L. Mcnaughton, Ivan Skelin. Distinct functional connectivity patterns predict the peri-ripple population activity in human amygdala and temporal cortex. Society for Neuroscience. Chicago IL, 2019
5. Shiting Ma, Michelle Paff, Ivan Skelin, **Haoxin Zhang**, Jack J. Lin; Changes in frequency power during anesthesia induction with propofol: A comparison between frontal cortex and the mesial temporal lobe. Society for Neuroscience. Chicago IL, 2019
6. **Haoxin Zhang**, Ivan Skelin, Shiting Ma, Bryce A. Mander, Jack J. Lin. Distinct high gamma activity profiles during REM and nREM sleep in the human hippocampus and neocortex. Human Single-Neuron Conference. Pasadena CA, 2018
7. **Haoxin Zhang**, Jie Zheng, Jack J. Lin. Medial temporal-medial prefrontal theta phase synchrony binds human associative memory. Society for Neuroscience. San Diego CA, 2018
8. Ivan Skelin, Jie Zheng, **Haoxin Zhang**, Bruce L. Mcnaughton, Jack J. Lin; Sharp wave ripple cross-structure phase locking to cortical oscillations predicts the modulation of cortical activity in human intracranial recordings. Society for Neuroscience. San Diego CA, 2018
9. **Haoxin Zhang**, Oliver Wang. Cortical processing of emotion: an electroencephalogram study. Society for Neuroscience. San Diego CA, 2016

10. **Haoxin Zhang**. Neural responses to emotion: An event-related potentials study. Society for Neuroscience. Chicago IL, 2015

PROFESSIONAL ASSOCIATIONS

Society for Neuroscience

Graduate Association of Biomedical Engineering Students in UC Irvine

ABSTRACT OF THE DISSERTATION

Emotional Processing and Pattern Separation in Medial Temporal Lobe

By

Haoxin Zhang

Doctor of Philosophy in Biomedical Engineering

University of California, Irvine, 2022

Professor Jack J. Lin, Chair

Remembering a salient experience and discriminating between similar events are two cognitive abilities that are critical for human adaptation and survival. They depend on optimal functioning of neurobiological mechanisms underlying episodic memory processing. Enhanced memory for emotional stimuli requires more pronounced memory encoding of emotional, relative to neutral stimuli. Discrimination between two different events depends on a neural computation of pattern separation. However, the exact neural mechanisms underlying memory encoding and pattern separation are still unclear.

In the present thesis, I applied an emotional memory encoding and discrimination task in human participants and examined the neural dynamics in the medial temporal lobe, known as a critical brain structure for memory processing. Local field potentials were recorded in the amygdala, hippocampus and peri-hippocampal regions, including entorhinal, parahippocampal and perirhinal cortex, from drug-resistant epilepsy patients undergoing pre-surgical

monitoring. First, I examined the role of hippocampal awake sharp-wave ripples (aSWRs) on emotional memory encoding, finding increased aSWRs after encoding of emotional, compared to neutral stimuli. The aSWRs were accompanied by memory reinstatement in the amygdala-hippocampal network. Additionally, the cross-structure joint-reinstatement during aSWR events was predictive of latter memory performance. Next, I investigated the neural mechanisms of pattern separation during memory discrimination. Stimulus-specific neural representation in the hippocampus predicted memory discrimination. Moreover, I designed a machine learning algorithm to decode the neural state, and found that the neural trajectory of dentate gyrus/Cornu Ammonis3 (DG/CA3) visited more unexplored state and predicted successful pattern separation. Finally, continuously increasing dimensionality predicts correct discrimination, suggesting code expansion as a mechanism implementing the pattern separation in the DG/CA3 region. Overall, these results suggest a putative mechanism for encoding and discriminating for emotional experience.

CHAPTER 1

Background and Significance

1.1 Different Memory System

"Memory is the scribe of the soul."

As quoted by Aristotle, memory is essential to the human expereience. Considerable research suggests the existence of partially distinct memory systems in the brain, including working, declarative and procedural memory (Squire, 1992; Lum et al., 2012). Working memory supports the maintenance and manipulation of ongoing information over a short time period, in the order of seconds, whereas declarative and procedural memory supports the long-term instantiation of knowledge for years (Lum et al., 2012). Declarative memory supports the encoding, storage, and retrieval of personal experience (episodic memory) and general knowledge about the world (semantic memory). Procedural memory, or non-declarative memory, encompasses a collection of nonconscious learning capacities which are expressed through performance but cannot be expressed verbally (Squire & Zola, 1996). The procedural system underlies a variety of perceptual, motor, and cognitive skills such as walking or riding a bicycle (Lum et al., 2012). The classification of declarative and non-declarative memory was historically introduced based on the converging

evidence from animal and human studies that the hippocampus and the medial temporal lobe (MTL) are essential for episodic memory formation. MTL lesions impair the ability to acquire new material (anterograde amnesia) and recall previously acquired information (retrograde amnesia) but preserve normal intellectual function (Squire 1992). While that the hippocampus and adjacent MTL structures are essential to memory processing, the exact mechanism is still unclear. The present thesis aims to investigate the neural dynamics underlying episodic memory formation and retrieval in the human medial temporal lobe.

1.2 Anatomical Overview of Medial Temporal Lobe

The medial temporal lobe consists of the hippocampus, amygdala, and their surrounding cortices, including the perirhinal cortex (PRC), parahippocampal cortex (PHC), and entorhinal cortex (EC). Information enters the MTL throughout different pathways from the neocortex. The perirhinal cortex receives inputs from the ventral visual stream area and has been shown to encode non-physical features of the object, such as the face (Miyashita, 2019). Animal and human neuroimaging studies show visual memory is more dependent on perirhinal cortex (Squire et al., 2004). The parahippocampal region is interconnected with multiple unimodal and multimodal cortical areas such as the temporal, parietal, and prefrontal cortex. It encodes a variety of aspects of the environment, such as spatial information (Aminoff et al., 2013). The entorhinal cortex receives input from both PRC and PHC, and projects to the hippocampus (Fig 1.1). The EC can be thought of as a gateway to the hippocampus for episodic memory formation, and its role on spatial and cognitive map is well established

(Fyhn et al., 2004; Tolman 1948; Epstein et al., 2017). The hippocampus is the ultimate recipient of convergent projections from the PRC, PHC and EC (Fig 1.1).

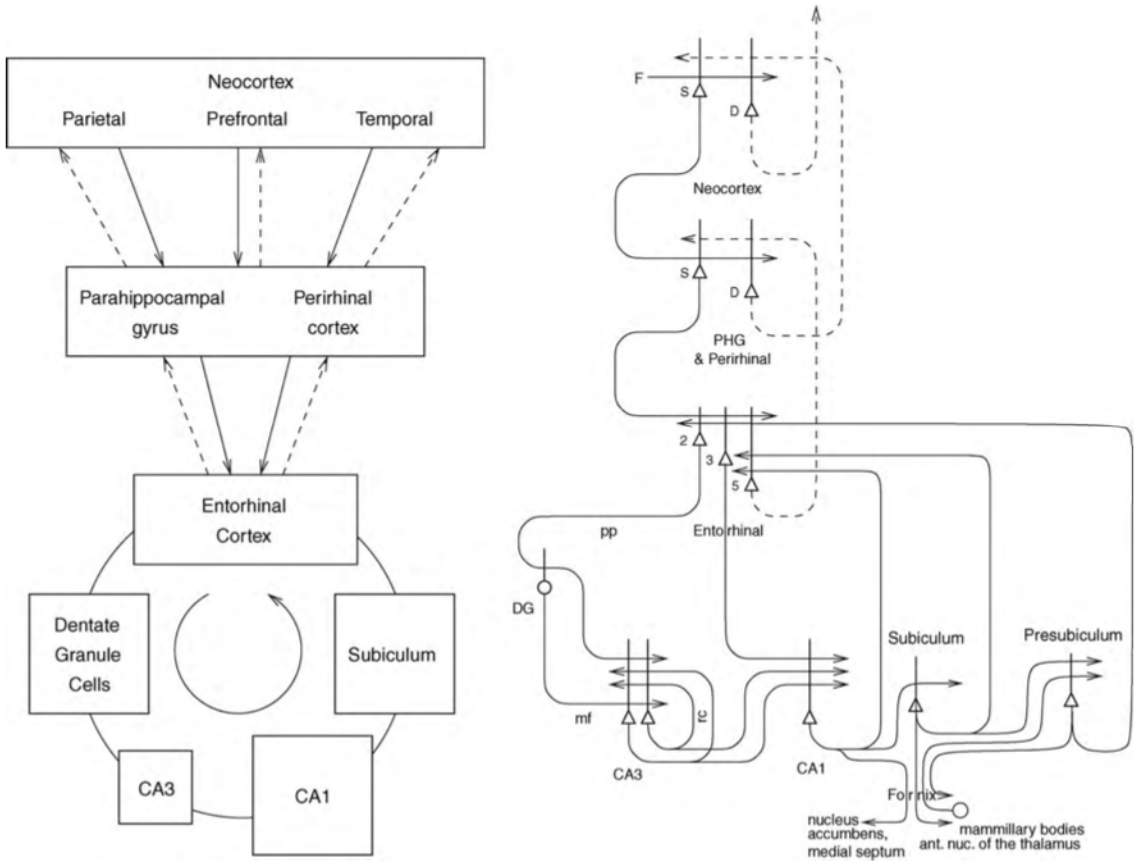


Fig 1.1 Schematic view of the anatomical circuits of the medial temporal lobe and its surrounding areas. Solid lines indicate forward connection from the associative cortex to the hippocampus, while dash lines indicate backward projection in the other direction. Abbreviations: DG: dentate granule cells, F: forward input from primary cortex, mf: mossy fibers, PHG: parahippocampal gyrus and perirhinal cortex, pp: perforant path, rc: recurrent

collateral, S: superficial pyramidal cells, 2: layer 2 of entorhinal cortex, 3: layer 3 of entorhinal cortex, 5: layer 5 of entorhinal cortex. (Figure adapted from Rolls, 2014)

The hippocampus consists of the Dentate Gyrus (DG), a series of Cornu Ammonis areas, and the subiculum (Fig 1.1). The trisynaptic circuit is the basic anatomical circuitry in the hippocampus, which involves three major cell groups: granule cells in the DG, pyramidal neurons in the Cornu Ammonis area 3 (CA3), and the Cornu Ammonis area 1 (CA1). The dentate granule cell receives input through the perforant path (pp) from layer 2 of EC (Witter 1993), and projects onto the pyramidal neuron in the CA3 region through mossy-fibres (mf). The CA3 pyramidal neuron receives three types of afferent connection: pp inputs from the EC, mf inputs from DG, and recurrent collaterals (rc) from other CA3 pyramidal neuron axion. Rodent studies show strong recurrent connections where each neuron roughly receives 12000 rc. The CA3-CA3 recurrent system is even more extensive in primates than in rats (Roll 2010; Kondo et al., 2009). It is followed by a moderate pp connection where each neuron receives 3600 pp inputs from EC. Each neuron receives roughly 46 mf connection resulting in sparse connectivity from dentate granule cells (Rolls et al.,1998; Treves et al., 1994; Rolls 2010). In turn, the CA3 pyramidal neuron project to CA1 neuron through Schaffer collaterals (sc). Additionally, layer 3 of the EC pyramidal neuron axon terminates in the CA1 (Fig 1.1).

1.3 Computation of Episodic Memory Processing

One fundamental feature of episodic memory is the associative binding of different elements of an episode (Davachi & Wagner, 2002). For example, a memory of riding a bicycle binds multiple aspects of the original event together, such as the location, the time, and the style of the bicycle. The collective combination of event details defines a unique memory representation for memory storage in the brain. This memory system is able to recover its original memory when presented with only partial information. For example, one might remind herself of riding a bicycle with her daughter in the park after seeing a similar bicycle in a different place.

Hebb and Marr proposed an auto-associative network as a general model of associative learning (Fig 1.2) that could support this remarkable ability to reconstruct past experiences. Associative learning occurs when multiple inputs are activated simultaneously, and the network should have the ability to associate the inputs together and retrieve their original pattern even when some of the inputs are absent. The proposed neural network consists of a population of the principal neurons. Each of the neuron's dendrites receives multiple synaptic inputs from pathways representing different aspects of the ongoing event, such that the input activation pattern represents the information of an event. Associative learning happens when the synaptic modification obeys the Hebbian learning rule that synapses are strengthened if two synaptic inputs are active in close millisecond proximity (Hebb, 1949). This encoding process ensures the uniqueness of each memory representation because different input patterns result in different synaptic connectivity patterns supporting distinct memories.

During retrieval, the network can recall the whole memory pattern by the mechanism of recurrent activation when only partial input patterns are presented, known as a process of pattern completion (Marr, 1971). From the dynamical system's perspective, the strengthened synaptic connections form a stable attractor in the state-space, where each neuron activation is treated as a hidden variable. During the retrieval, the input response of the dynamical system may fall into the attractor (i.e., the encoded memory) if the input pattern (i.e., cue) is partially similar to the original input.

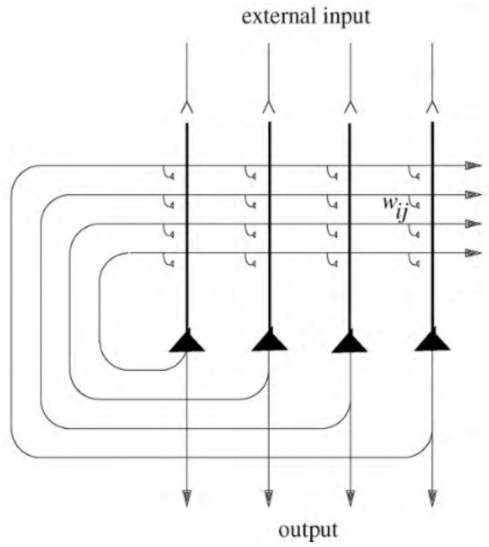


Fig. 1.2 Circuit diagram of an auto-associative network.

Each principal neuron receives external inputs as well as recurrent connection from other neuron. Solid triangle indicates principal neuron.

Experimental studies have reported neurobiological correlates of core elements of the model. The Hebbian rule of synaptic strengthening is analogous to long-term potentiation (LTP), where synaptic connections strengthen if post-synaptic firing occurs right after pre-synaptic firing. An auto-associative network architecture has been found in the CA3 region of the hippocampus, where the CA3 principal neuron receives extensive projection from other principal neurons in the CA3 through recurrent collaterals. Moreover, the

hippocampus is located at the top of the anatomical hierarchy, where it receives direct and indirect inputs from the primary and associative cortex. This allows the formation of association between inputs originating from very different parts of the cerebral (Rolls 2010). For example, a memory of riding a bicycle requires the network to associate the visual appearance of the bicycle originated from the visual cortex with the emotional perception of the experience originated from the amygdala.

1.4 Pattern Completion and Pattern Separation

In the last section, I discussed the auto-associative network as a general model for episodic memory. During memory formation, it activates an ensemble of interconnected neurons and induces synaptic potentiation in the corresponding recurrent synapses (Guzman et al., 2016). During retrieval, a partial pattern initially activates a subset of the ensemble, but the network subsequently recruits the remaining cells via a strong recurrent connection (Guzman et al., 2016). This retrieval process can be viewed from the dynamical system point of view where neuronal activation vector is defined as hidden state variables. A partial input may drive the neural state trajectory to fall into an attractor state originally formed during encoding (Fig. 1.3). This process is termed pattern completion and is proposed as a plausible neural computation occurring in the hippocampus supporting memory recall.

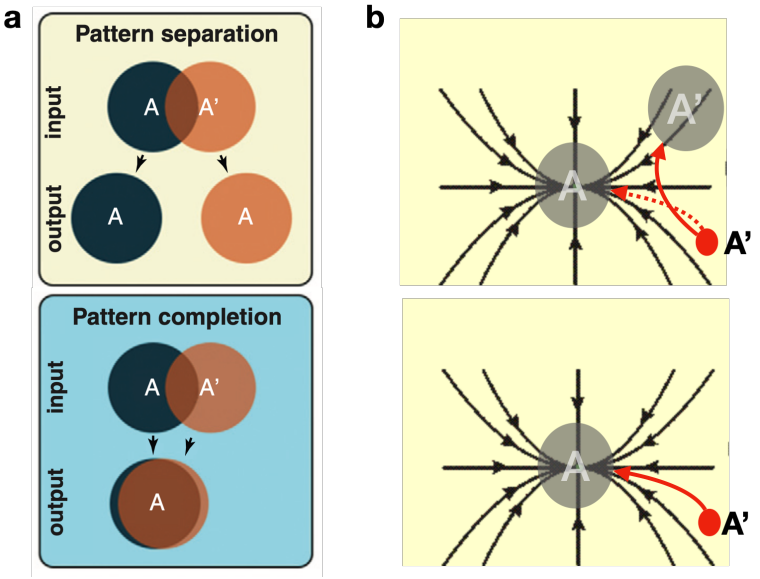


Fig. 1.3 Pattern separation and pattern completion. a, schematic view of two competing process. b, neural state trajectory undergoing pattern separation (upper) and completion (lower) in the dynamical system.

However, pattern completion alone does not allow the memory system to discriminate a new but similar memory from an old one. Hebb-Mar associative learning theory suggests that to maximize information storage with minimal interference, the hippocampal associative network performs two competing, yet complementary, processes: pattern completion and pattern separation (Guzowski et al., 2004; McNaughton and Morris, 1987; Rolls and Treves, 1998; Neunuebel and Knierim, 2014). Pattern separation is a neural computation by which the network reduces the overlap between similar input patterns leading to non-overlapping patterns (Fig. 1.3).

Although the exact biological implementation of pattern separation is still unclear, computational studies propose serval potential mechanisms. David Albus proposed a mechanism of expansion recoding in a feedforward network where the first layer inputs onto a significantly larger population in the second

layer (Albus, 1971). This process allows a downstream decoder to linearly classify two similar inputs (Cayco-Gajic et al., 2019). Consistent with this idea, the entorhinal projects onto a much denser population of DG granule cells (Yassa and Stark, 2011; Knierim and Neunuebel, 2016; Rolls, 2016). Moreover, the remapping in the DG is more pronounced than CA3 and CA1 regions (Leutgeb et al., 2007; Neunuebel et al., 2013; Neunuebel and Knierim, 2014).

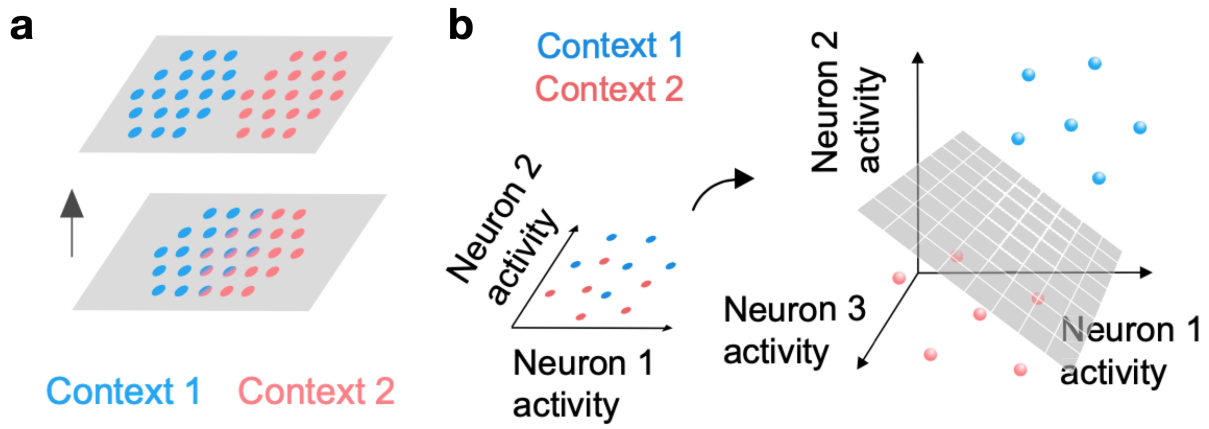


Fig 1.4 High dimensional neural coding space support pattern separation

a. Overlapping representation become distinctive representation under pattern separation process. **b,** Albus suggested that expending the dimensionality in the neural state space enable the downstream region to classify two stimuli (Albus, 1971). (Adapted from Cayco-Gajic et al., 2019)

An important concept needed to describe this mechanism is the dimensionality of a system, which refers to the number of independent variables needed to describe a system's all possible states. In the neural network system, the dimensionality size is the number of neurons and the activity of an individual neuron is an independent variable, with a neuronal ensemble forming a neural state space. The coding subspace is derived from the neural activity space. It comprises the full set of population representations of all possible stimuli (Druckmann and Chklovskii, 2012; Cayco-Gajic et al., 2019). Higher-dimensional coding subspaces facilitate the separation of overlapping activity patterns by providing a larger activity space wherein representations are embedded, thereby increasing their linear separability (Cover, 1965; Cayco-Gajic et al., 2019; Fig. 1.4). In contrast, low-dimensional coding subspaces limit the extent of possible representations that can be generated. However, direct evidence of increased dimensionality as a mechanism of pattern separation has not been shown in humans.

1.5 Emotional Modulation of Episodic Memory

Emotions can strengthen memories: we remember events that carried substantial positive or negative emotions better and more vividly than neutral experiences. (Kensinger 2009; Cahill & McGaugh, 1998) The memory advantages of emotional experience appear immediately after encoding and become more

pronounced over time (Yonelinas & Ritchey, M., 2015). Multiple neural mechanisms have been proposed to support this behavioral observation. The modulation theory suggests that the Glutamate Amplifies Noradrenergic Effects (GANE) during the encoding may prioritize preferred memory storage for emotional memories. Furthermore, the modulation theory proposes that the secretion of certain hormones such as adrenaline and cortisol during encoding leads to a greater extent of memory consolidation during sleep, which explains the enhancement effect with a time delay. (Cahill & McGaugh, 1998; Talmi, 2013; Yonelinas & Ritchey, M., 2015)

Sleep plays a critical role in memory enhancement. Behavioral studies show that sleep deprivation impairs post-learning memory performance. More specifically, selective disruption of hippocampal sharp-wave ripples (SWRs) impairs post-sleep memory performance, whereas prolonging SWRs enhances memory (Buzsáki, 2015). The SWR is a transient high-frequency oscillation (80-150 Hz) that appears in the hippocampal local field potential (LFP) and is accompanied by synchronized neuronal population activity (Logothetis et al., 2012; Buzsáki, 2015). Recent studies show anatomical distributed neocortical activation coincides with hippocampal SWR (Skelin et al., 2021; Ramirez-Villegas et al., 2015). From the perspective of Hebb's associative learning model, synchronized reactivation of encoded memory patterns strengthens synaptic connectivity, stabilizing memory representation (Fries, 2005). Hippocampal reactivation during SWR has been proposed as a neural substrate enabling the

binding of multiple memory traces across structures through LTP induction (Buzsáki, 2015).

Although behavioral studies reveal a critical role of sleep on remembering an emotional experience, the exact neural mechanism is largely unexplored. A recent rodent study highlights the role of hippocampal SWR in reactivating fearful memory engram in the amygdala-hippocampal network during the NREM sleep as a potential mechanism for consolidation (Girardeau et al., 2017).

SWRs have also been observed during immobile wakefulness (Carr et al., 2011). The awake state occurrence of SWRs is more prevalent after exposure to salient or reward-associated context and is associated with hippocampal place cell reactivation (Joo & Frank 2018). Significantly, the reactivation of rewarded or fearful memory is enhanced relative to neutral memories. However, whether awake SWR (aSWR) plays a role in emotional memory processing and supports beneficial memory storage is unknown. In the present thesis, I hypothesized that the aSWR event provides a neural substrate for the coordinated reactivation of the memory engram in the amygdala-hippocampal network enabling enhancement of memory for an emotional experience.

1.6 Human Electrophysiological Recording

Among the range variety of brain imaging techniques employed in modern neuroscience research, there are four main types of functional recording methods available in human experiments: scalp electroencephalogram (scalp EEG), intracranial electroencephalogram including electrocorticography (ECoG) and

stereoencephalography (SEEG), magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI). Intracranial EEG (iEEG) is an invasive electrophysiological recording approach that monitors extracellular electric field within a volume of brain tissue by surgically implanting either ECoG or depth electrodes (SEEG) targeting deep brain structures such as the MTL (Buzsaki et al., 2012). This approach characterizes sub-millisecond time resolution not available with fMRI, as well as sub-millimeter spatial resolution, not available with scalp EEG (Slutzky et al., 2008; Buzsaki et al., 2012). The obtained local field potential (LFP) reflects the ensemble activity of tens of thousands of nerve cells (Canolty and Knight, 2010). In the present thesis, we recorded LFP on pre-surgical epilepsy patients with depth electrodes implanted in the MTL, providing a rare opportunity to study neural dynamics with superb temporal and spatial resolution.

1.7 High Frequency Activity and Memory Processing

1.7.1 Rhythmic Activity underlying Memory Processing

Rhythmic activity is the most prominent feature observed in the local field potential. Oscillatory activity in humans was discovered by Hans Berger using scalp EEG recordings where he described as an 8 to 12Hz rhythm. Rhythmic activity in distinct frequency bands characterize changes in response to the sensory, motor, and cognitive events (Canolty & Knight, 2010). Among activities in multiple bands, theta oscillation (4-12Hz) and gamma activity (30-250Hz) are

both associated with declarative memory processing (Lisman & Jensen, 2013). Extensive animal and human studies suggest that successful memory formation and retrieval are associated with prominent theta and gamma power, as well as strong coupling between the two activities (Mormann et al., 2005; Montgomery and Buzsáki,2007; Sauseng et al., 2009; Shirvalkar et al., 2010; Axmacher et al., 2010; Carr et al., 2012; Zheng et al. 2017&2019; Stevenson et al. 2018). More specific, the phase of theta rhythms modulates the amplitude of gamma frequency activity. In other words, the gamma activity power tends to increase or decrease at a particular phase of theta rhythm. A similar phenomenon, known as phase precession, was observed that the place cell representing different spatial locations tends to fire at distinctive phases of underlying theta oscillation using single-unit recording in rodents (Mehta et al., 2002; Battaglia et al., 2004; Lisman & Jensen, 2013). Although the exact mechanism of generating theta-gamma coupling is unclear, both the field potential and single-unit recording studies support a theta-gamma neural code theory (Lisman & Jensen, 2013). The theory proposes that the subset of cells that fire during a given gamma cycle (sometimes referred to as a cell assembly or an ensemble) form a spatial pattern that represents a given item so that largely non-overlapping assemblies representing different items are active at different theta phases (Lisman & Buzsáki, 2008; Lisman & Jensen, 2013).

Recent studies report a specific type of gamma oscillation, which occurs in different frequency ranges with short temporal duration, plays a role in episodic

memory processing and maintains multiple items in the working memory (Sirota, Anton, et al.,2008; Lopes-dos-Santos et al. 2018; Lundqvist et al. 2016&2018; Kamiński et al. 2017; Navas-Olive, Andrea, et al.,2020). Following the theta-gamma neural code theory, each burst of gamma oscillatory activity reflects distinctive cell assemblies firing, which may represent different items. A series of theoretical works describe a potential mechanism of frequency-specific gamma bursting activity. In the next section, we first briefly describe the biophysical mechanism of gamma oscillation as well as theta-gamma coupling. We then present computational modeling research on working memory supported by gamma bursting activity.

1.7.2 Theoretical Work of Frequency-specific Gamma Bursting Activity

Theory proposes that two elementary circuit architectures generate stable gamma oscillations with both utilizing inhibitory interneurons (Buzsáki et al. 2012).

In the first scenario, when the interneurons are reciprocally connected, the synchronization starts when a subset of the interneurons synchronously discharge by chance, providing inhibitory input to other interneurons, until the decay of GABA-mediated hyperpolarization. Regardless of whether the initial input is tonic or stochastic (Fig 1.5a-b), the system starts to synchronize at the frequency of gamma determined mainly by the time constant of the IPSP and the net excitation of the interneurons.(Whittington et al. 1995, Wang & Buzsaki 1996 ,

Buzsáki et al. 2012). In the second scenario, where excitatory pyramidal and inhibitory neurons are reciprocal connected (Fig. 1.5c), the synchronization is caused by the fast excitation and delayed feedback inhibition of the pyramidal neuron. The amount of delayed time of inhibition determines the frequency of gamma oscillation.

Both scenarios could cooperate to generate gamma oscillation with different frequency characteristics depending on the underlying neuronal connection, input excitation and synaptic properties. For example, Rennó-Costa (Rennó-Costa et al. 2019) shows different feedforward inhibition would alter the frequency profile of the gamma oscillation in an excitatory-inhibitory (E-I) neuronal population (Fig 1.5d-e; Rennó-Costa et al. 2019)

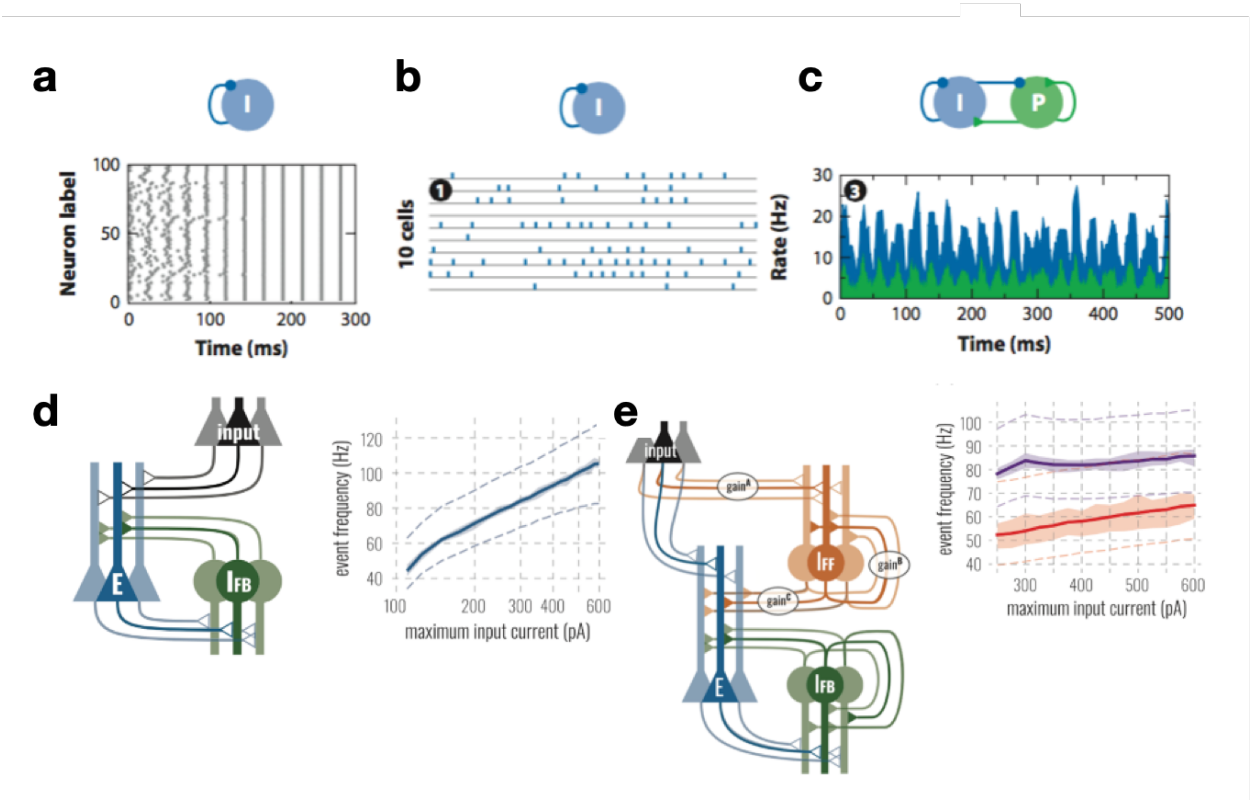


Fig. 1.5. Model of Gamma Oscillation (a-c) demonstrates two types of elementary circuits generating gamma oscillation. **a,** I-I circuit with tonic input generate periodically firing of interneuron at 40Hz; **b,** I-I circuit with stochastic input generates sparsely synchronous oscillation; **c,** Reciprocally E-I model where pyramidal neuron(P) send fast excitation via AMPA receptor to interneuron which feedback inhibit P via GABA receptor. (Adapted from Buzsáki et al. 2012) (d,e) shows an example of an elementary model cooperating together to generates complex frequency profile. **d,** The reciprocally connected population of excitatory (E) and inhibitory population (IFB) generates gamma oscillation with a varied frequency response corresponding to varied input excitation onto excitatory neuron.; **e,** An I-I circuit generates inhibition onto E-I circuit. By introducing a self-connected inhibitory network IFF and further feedforward inhibit the excitation of E, the center gamma frequency of E-IFB network can be stabilized and varied with the strength of feedforward inhibition. By increasing the $gain^c$, the gamma frequency of E-I circuits decreases.(red: $gain^c$, = 20% and $gain^A$, = 200%;purple: $gain^c$, = 40% and $gain^A$, = 80%; $gain^B$, = 200% for both ; Adapted from Rennó-Costa et al. 2019)

In principle, the generation of theta-gamma coupling requires 1) the generation of theta/gamma oscillation; and 2) the coupling between two oscillators (Hyafil et al. 2015). The two oscillatory circuits can either be independently generating theta and gamma rhythm and coupling through a unidirectional or bidirectional connection or sharing a common subpopulation

(Fig 1.6). In the abovementioned example (Rennó-Costa et al. 2019), the theta-gamma coupling arises by adding unidirectional theta rhythmic input into the E-I circuit, and the gamma oscillation shows the highest amplitude in the peak phase of theta oscillation (Fig 1.6).

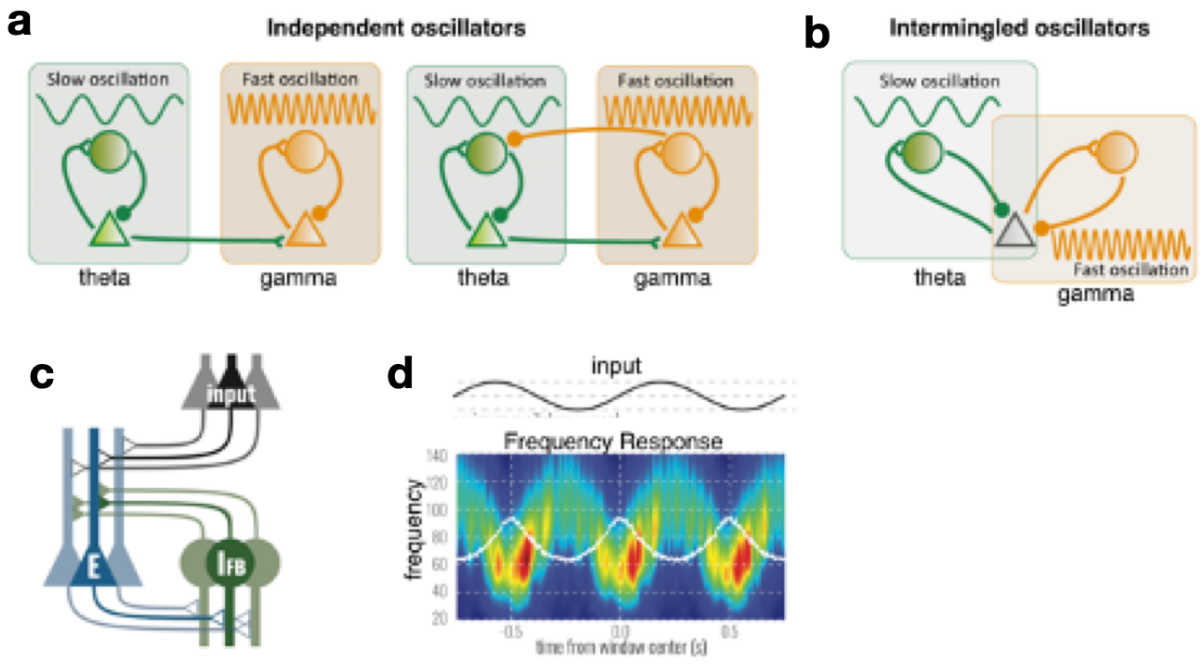


Fig. 1.6 Model of Theta-Gamma Coupling a&b demonstrate two regimes of coupling of fast oscillation and slow oscillation could occur. **a,** Independent oscillator coupling through unidirectional (left) or bidirectional (right) connection of two regions; **b,** Intermingled oscillator share a common subpopulation of neurons within the same region. (Adapted from Hyafil et al. 2015) **c,d,** the E-I circuit with theta rhythm input generate coupled gamma oscillation. The right bottom figure is the averaged power frequency plot at each phase of the theta oscillation (white line). x-axis: phase of theta, y-axis: frequency of gamma oscillation, color: power of gamma oscillation. (Adapted from Rennó-Costa et al. 2019)

Encoding and maintaining multiple items in working memory is associated with theta/gamma power increase and theta-gamma coupling (Axmacher et al. 2010; Tort et al. 2009; Fuentemilla et al. 2010). Moreover, item-specific gamma oscillation is time-locked to a particular phase of the theta cycle (Heusser et al. 2016; Lundqvist et al. 2016&2018). A series of biophysically

detailed models were built to study the neuronal and LFP dynamics underlying working memory.

Lundqvist et al. built a model with neocortical network architecture, containing nine hypercolumns where each hypercolumn containing 49 non-overlapping minicolumns including excitatory pyramidal cells and two types of inhibitory interneurons, i.e., regular spiking non-pyramidal cells (RSNP) and basket cells (Lundqvist et al.2006,2010,2011; Herman et al. 2013). Pyramidal cells sharing the same minicolumn have dense (25%) recurrent connections and are viewed as the basic computational units of the network. However, attractors representing different items are stored by means of sparse long-range connectivity between minicolumns from separate hypercolumns. Each hypercolumn acts as a local winner-take-all module using basket cell feedback inhibition. That is, the network consists of several connected winner-take-all modules. Another set of interneurons, RSNP cells, provided long-range synaptic inhibition supporting competition between attractors (Lundqvist et al., 2011). To conclude, each burst of frequency-specific gamma oscillatory activity reflects distinctive cell assemblies representing different items, and the frequency characteristic is defined by its underlying synaptic properties generating gamma oscillation.

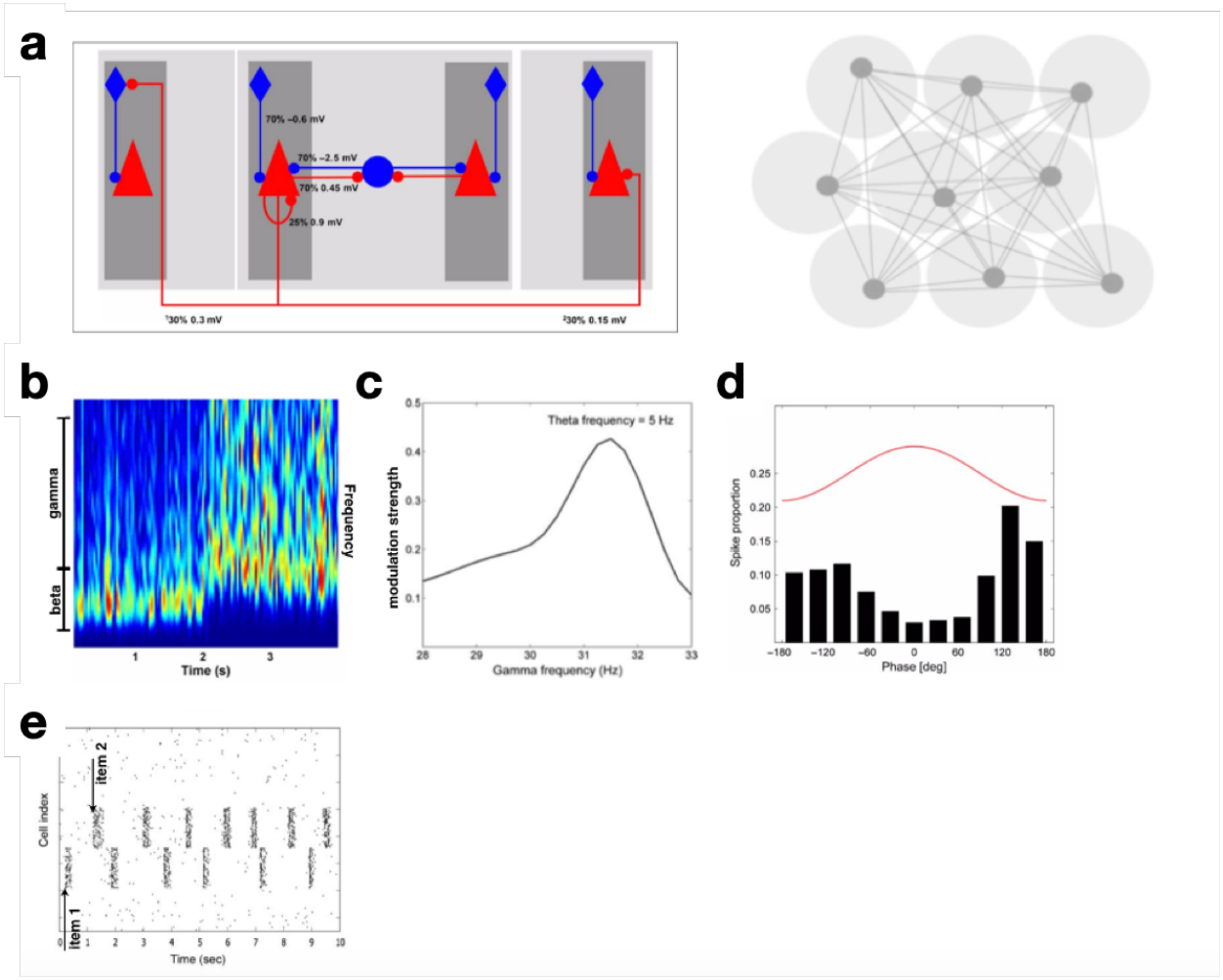


Fig.1.7 Computational Model of theta-Gamma Coupling during Working Memory
a, Schematic drawing of the model structure. Light squares mark hypercolumns, and dark squares mark minicolumns. Red triangles represent pyramidal cells, blue circles represent basket cells, and blue rhombs represent RSNP cells. Numbers show the likelihood of a connection between prepopulation and postpopulation and effect of a synaptic event in soma of the postsynaptic cell. (Adapted from Lundqvist et al. 2010) **b,** The time-frequency plot shows two states: the ground state from 0 to 2 sec and the active state from 2 to 4 sec. (Adapted from Lundqvist et al. 2010) **c,** Modulation strength with theta oscillation (5Hz) shows a peak at ~30Hz. **d,** Distribution of spiking probability at different phases of theta oscillation. **e,** Raster plot of a subset of simulated population activity when two items were loaded. Cell assemblies representing two items fire alternatively. (Adapted from Lundqvist et al. 2011)

The network could either be in a state where one of the stored attractors was active or in a ground state where all minicolumns exhibited low-level activity (Amit & Brunel, 1997; Lundqvist et al. 2011). The ground state simulates LFP mainly exhibits oscillation in beta frequency, mainly driven by the firing of

RSNP cell, while the active state exhibits increased theta and gamma activity (Fig. 1.7.b&c). Notably, the gamma rhythm was generated by the E-I circuits of pyramidal cell and basket cell within the hypercolumn with the frequency defined by the synaptic properties between them. The model showed that each frequency-specific gamma oscillation represented different items in the working memory. Moreover, it occurred in the bursting regime where each gamma oscillation existed for a short duration due to the RSNP's long-range inhibition, so that it supported the transition between different ensembles. Consistent with the experimental findings, both gamma and neuronal spiking activity tend to lock to a certain phase of underlying theta oscillation (Lundqvist et al., 2011).

In the present thesis, I examined the frequency profile in the gamma range (30-280Hz, High Frequency Activity) to investigate the cell assemblies activation representing different aspects of memory in the human MTL.

CHAPTER 2

Neural Dynamics of Emotional Memory Processing in the Amygdala-hippocampal network

2.1 Abstract

Intracranial recordings from the human amygdala and the hippocampus during an emotional memory encoding and discrimination task reveal increased awake sharp-wave/ripples (aSWR) after encoding of emotional compared to neutral stimuli. Further, post-encoding aSWR-locked memory reinstatement in the amygdala and the hippocampus was predictive of later memory discrimination. These findings provide electrophysiological evidence that post-encoding aSWRs enhance memory for emotional events.

2.2 Introduction

Multiple mechanisms have been proposed to explain the prioritized encoding of emotional experiences (Cahill and McGaugh, 1998; Kensinger ,2009; Szőllősi and Racsmány, 2020), including the neuromodulatory effects on plasticity and the interplay between the amygdala and the hippocampus (Cahill and McGaugh, 1998; Talmi 2013; Yonelinas and Ritchey, 2015). Several studies have found memory reinstatement during the immediate post-encoding period to be predictive of later memory performance (Ben-Yakov and Henson, 2018 ; Sols et

al., 2017). Sharp-wave/ripples (SWRs) are transient hippocampal oscillations (80-150 Hz), associated with synchronous neural activation in the hippocampus and the amygdala (Logothetis et al., 2012; Skelin et al., 2021), and are implicated in the binding of anatomically distributed memory traces (Buzsáki ,2015). Behaviorally relevant reactivation of emotional memory occurs during aSWRs (Wu et al., 2017), and disruptions of post-experience aSWR interfere with memory utilization (Jadhav et al., 2016). Based on these findings, we hypothesized that aSWRs occurring immediately after stimulus encoding (post-encoding) facilitate emotional memory discrimination through the coordinated hippocampal-amygdala memory reinstatement. Using intracranial electroencephalographic (iEEG) recordings in epilepsy patients during the performance of an emotional encoding and discrimination task, we first confirm reports of better discrimination memory for arousing stimuli (Szőllősi and Racsmány, 2020). Next, we demonstrate that the number of aSWR events immediately after encoding is associated with both stimulus-induced arousal and the accuracy of later discrimination. Finally, the coordinated memory reinstatement between the amygdala and the hippocampus during post-encoding aSWRs is predictive of later memory discrimination performance, with the amygdala reinstatement showing a directional influence on the hippocampal reinstatement. Together, these findings provide evidence that aSWRs-mediated memory reinstatement in the amygdala and hippocampus as a mechanism accounting for better remembering of emotional experiences.

2.3 Materials and Methods

Subjects

Intracranial electroencephalography (iEEG) recordings were obtained from 7 subjects (3 females; mean age ± SD = 33 ± 16), undergoing presurgical monitoring of epileptic foci at the University of California Irvine Medical Center (UCIMC) Epilepsy Monitoring Unit. The individual subject demographic information is shown in Table 1. Only the subjects with the correct discrimination rate of Novel trials >= 85% (see Emotional memory encoding and discrimination task) were included in the analysis. Electrode placements were determined entirely based on clinical considerations. All the research procedures were approved by the UCI Institutional Review Board and data was collected following informed consent.

Statistics

All the statistical tests were performed with the individual subject as the unit of analysis. Unless stated otherwise, all the parametric statistical tests (e.g., Wilcoxon signed-rank test, t-test) were two-tailed. The effects of valence, stimulus-induced arousal and similarity on stimulus discrimination (Fig. 1c) were assessed using the logistic linear mixed-effect model (for details, see Behavioral Analysis). Conditional comparisons of aSWR occurrence (correct/incorrect discrimination or high/low arousal; Fig. 2.2c) were done using the Wilcoxon signed rank test ($p < 0.05$). Statistical significance of aSWR-locked memory reinstatement strength (Fig. 2.3b) was assessed by comparing the real

test statistics with empirical null distribution, obtained using Monte Carlo method (for details, see Representational Similarity Analysis). We implemented the cluster-based nonparametric permutation test(Maris & Oostenveld, 2007) to assess the conditional differences (correct/incorrect discrimination or high/low arousal) of memory reinstatement strength (Fig. 2.3c), mutual information (Fig. 2.3e), by randomly shuffling the conditional trial labels 1000 times (for details, see Representational Similarity Analysis).Similarly, the significant temporal windows for the cross structure aSWR-locked joint memory reinstatement (Fig. 2.3d) were assessed by comparing to empirical null distribution (for details, see Joint-reinstatement Analysis).

Emotional memory encoding and discrimination task

The emotional memory encoding and discrimination (EMOP) task consists of encoding and discrimination blocks. During the encoding block (148 trials), each trial consists of a cross fixation (1000 msec), followed by stimulus encoding (2000 msec) and self-paced post-encoding response period (up to 2000 msec). During the post-encoding response period, subjects are asked to classify the stimulus emotional valence as either negative, neutral or positive, using the corresponding laptop key. During the retrieval block (290 trials), trial time structure is identical to encoding phase. Following the cross fixation (1000 msec), the subjects are presented for 2000 msec with a stimulus identical (Repeat, 54 trials), slightly different (Lure, 97 trials) or unrelated (Novel, 139 trials) to previously encoded stimuli. Next, during the self-paced memory discrimination

epoch (up to 2000 msec), subjects are asked to discriminate if the presented stimulus was seen during encoding (Old) or not (New). Correct discrimination is defined as classifying the Repeat stimuli as Old and Lure or Novel stimuli as New. The stimuli were selected from the continuous distributions across the valence and stimulus-induced arousal axes (Supplemental Fig. 1). The same set of stimuli was used across subjects. In addition, the valence, arousal and similarity of each stimulus were rated by separate cohorts of healthy subjects. Specifically, a first cohort (N = 50, 32 females; age mean ± SD = 22 ± 5) rated the stimulus emotional valence on a continuous scale (range 1-9, with 1 denoting the most negative, 9 the most positive, and 5 neutral valence). Stimuli were assigned in Negative (valence <= 3.5), Neutral (3.5 > valence < 6) or Positive (valence >= 6) groups. Another cohort of healthy subjects (N = 16, 4 females; age mean ± SD = 23 ± 5) rated the stimulus-induced emotional arousal on a scale 1 - 9 (1 being the least and 9 being the most arousing). Finally, a third cohort (N = 17, 11 females; age mean ± SD = 20 ± 1) examined relative similarity on the scale 1-8 (Leal et al., 2014). The high correspondence of stimulus valence ratings obtained from study subjects and healthy population (match rate = 85.3 ± 1.3%) suggests the intact emotional processing in study subjects (Supplemental Fig. 1).

Behavioral Analyses

To assess the effects of valence, stimulus-induced arousal and similarity on Lure stimulus discrimination, we implemented the logistic linear mixed-effect model

$$y = \beta X + uZ + \varepsilon.$$

In this model, y indicates the responses across the individual Lure discrimination trials (0-Old; 1-New), $X = [x_1, x_2, x_3]^T$ denotes three fixed effect regressors (encoded stimulus valence and arousal as well as similarity between the encoded and Lure stimulus), $Z = [z_1]^T$ denotes random effect regressor (subject identity), β and u denote the fixed and random-effect regression coefficients, and ε denotes the error term. The model includes random intercept to incorporate individual subject differences. We normalized the valence, stimulus-induced arousal and similarity values relative to the scale of 0 to 1. The statistics reported in Fig. 1c corresponds to the fixed-effect coefficients β.

Data collection

The behavioral experiment was administered using the PsychoPy2 software(Peirce, 2009) (Version 1.82.01). The laptop was placed at a comfortable distance in front of the subject. The iEEG signal was recorded using a Nihon Kohen system (256 channel amplifier, model JE120A), with an analog high-pass filter (0.01 Hz cutoff frequency) and sampling frequency 5000 Hz.

Electrode localization

We localized each electrode using pre-implantation structural T1-weighted MRI scans (pre-MRI) and post-implantation MRI scans (post-MRI) or CT scans (post-CT). Specifically, we co-registered pre-MRI and post-MRI (or post-CT)

scans by means of a rigid body transformation parametrized with three translation in x,y,z directions as well as three rotations using Advanced Normalization Tools (ANTs https://stnava.github.io/ANTs/). We implemented a high-resolution anatomical template with the label of medial temporal lobe subfields(Leal et al., 2014) to guide the localization for individual electrodes. We resampled the template with 1mm isotropic, and aligned it to pre-MRI by ANTs Symmetric Normalization (Avants et al., 2011) to produce a subject-specific template. The electrode localization was identified by comparing the subject-specific template subfield area with electrode artifacts.(Fig. 2.2a) The localization results were further reviewed by the neurologist (J.J.L.).

Preprocessing

The signal preprocessing was done using the custom-written MATLAB code (Version 9.7) and Fieldtrip Toolbox(Oostenveld et al., 2011). The 60 Hz line noise and its harmonics were removed using a finite impulse response (FIR) notch filter (ft_preprocessing.m function in FieldTrip). The EEG signal was down-sampled to 2000 Hz, demeaned and high-passed filtered (cutoff frequency 0.3 Hz). The power spectrum density (PSD) was computed using the multitaper method with the Hanning window (ft_freqanalysis.m function in FieldTrip). All the channels were re-referenced to the nearest white matter channel from the same depth electrode, based on the electrode localization results. The interictal epilectic discharges were manually marked by an epileptologist (J.J.L.), using the ft_databrowser.m function in FieldTrip. The channels with severe contamination

and trials containing epileptiform discharges were excluded from further analyses.

Awake sharp wave ripple detection

Following the removal of channels with excessive epileptic activity and individual trials containing visually identified interictal epilectic discharges, awake sharp-wave/ripples (aSWRs) were detected on the remaining hippocampal channels, using the Freely Moving Animal Toolbox (FMA; http://fmatoolbox.sourceforge.net/). First, the iEEG traces from the trials used in the analysis were concatenated. Next, concatenated traces were bandpass-filtered (80 - 150 Hz, Chebyshev 4th order filter, function filtfilt.m in Matlab) and the voltage values during periods ± 75 msec around the trial onsets/offsets were set to zero, to avoid the edge effects resulting from filtering discontinuous traces. The analytical amplitude was obtained by computing the absolute value of Hilbert-transformed filtered trace (function hilbert.m in Matlab) and z-scored (Supplemental Fig. 2a). Detected events were considered aSWRs if the z-scored analytical amplitude remained above the lower threshold ($z = 2$) for 20 - 100 msec and if the peak value during this period exceeded higher threshold ($z = 5$). Only the channels with >150 detected aSWR events were used in the analysis. If the multiple channels from a single subject passed this criteria, a channel with highest number of detected aSWRs was selected for further aSWR-related analysis. Due to the low number of detected aSWRs, one subject was eliminated from the aSWR-related analysis.

Unsupervised decomposition of iEEG signal

To assess the memory reinstatement, high-frequency activity (HFA; 30-280 Hz) was used as an indirect measure of local populational activity(Canolty & Knight, 2010; Lopes-dos-Santos et al., 2018; Ray & Maunsell, 2011; Wixted et al., 2014). To avoid the effect of low-frequency harmonics on the HFA estimate, we applied the Ensemble Empirical Mode Decomposition(Lopes-dos-Santos et al., 2018; Wu & Huang, 2009) (EEMD; https://github.com/leeneil/eemd-matlab.git). Briefly, the EEMD decomposes a non-stationary signal into its elementary components, referred to as intrinsic mode functions(Wu & Huang, 2009) (IMFs; Supplemental Fig. 6). The procedure iteratively applies an empirical mode decomposition algorithm, while adding white noise to prevent the mode mixing(Huang et al., 1998; Wu & Huang, 2009). Using this approach, decomposition output entirely depends on the signal's intrinsic properties, avoiding prior assumptions(Huang et al., 1998; Lopes-dos-Santos et al., 2018; Wu & Huang, 2009). The resulting IMFs captured several canonical spectral features consistently across subjects and anatomical structures (Supplemental Table 2). Finally, the HFA time-series on individual channels were reconstructed by summing the channel-specific IMFs with center frequencies > 30 Hz(Lopes-dos-Santos et al., 2018).

Time-frequency representation of the HFA

The instantaneous spectral power at each time-frequency bin was derived from the reconstructed HFA time series (x), using a wavelet transform(Bârzan, 2020; Moca et al., 2021). This approach consists of convolving the time series $x$with a set of Morlet wavelets, parametrized by a range of cycle numbers (n = 2, 3, …, 10) at a given frequency f,

$$P_{f,n}(t) = \left|\psi_{f,n} * x(t)\right|, n = 2,3, \dots ,10$$

with $\psi_{f,n}$defined as

$$\psi_{f,n} = \frac{1}{B_n\sqrt{2\pi}} e^{-\frac{t^2}{2B_n^2}} e^{j2\pi ft}, where\ B_n = \frac{n}{5f}$$

and computing the geometric average $\widehat{(P}(f,t))$of resulting spectral power at each time- frequency bin:

$$\hat{P}(f,t) = \sqrt[9]{\prod_{n=2}^{10} \quad P_{f,n}(t)}\ .$$

This approach results in a high temporal and frequency resolution, facilitating the detection of narrow-band, transient oscillatory events(Bârzan, 2020; Moca et al., 2021). The wavelet center frequencies were within 30 - 280 Hz range, with 1 Hz increments. The wavelet cycle number range (2-10) is commonly used(Cohen,

2014). To avoid the edge effects, this procedure was applied on the entire individual recording sessions, and the resulting time-frequency response matrices were segmented into trial epochs (starting -1000 msec prior to stimulus onset and ending 1000 msec after the response time). The power within each trial epoch was then normalized by z-transforming each frequency bin and subtracting the average pre-trial baseline (-1000 - 0 msec, relative to stimulus onset(Cohen, 2014)).

Representational Similarity Analysis (RSA)

The representational similarity was quantified as the Spearman correlation between the HFA power spectral vectors (PSVs), for each combination of the encoding-response time bins from the same trial(Lohnas et al., 2018; Norman et al., 2019; Yaffe et al., 2014; Zhang et al., 2018) (Supplemental Fig. 6). Specifically, the instantaneous spectral power at each frequency was estimated for 100 msec time bins (10 msec step size, 90% overlap), producing the time bin - specific power spectrum vectors (PSV), spanning the encoding (2 sec time window after stimulus onset) and post-encoding response (time window after stimulus offset and before button press) periods:

$$\overrightarrow{PSV}_{encoding}(t_1) = \left[z_1(t_1), \ldots, z_{n_f}(t_1)\right]_{encoding}$$

$$\overrightarrow{PSV}_{response}(t_2) = \left[z_1(t_2), \ldots, z_{n_f}(t_2)\right]_{response}$$

Similar to previous studies(Lohnas et al., 2018; Norman et al., 2019; Pacheco Estefan et al., 2019; Staresina et al., 2016; Yaffe et al., 2014; Zhang et al., 2018), we computed Spearman's correlation as a measure of PSV similarity between the encoding time t_1 and response time t_2 for each encoded stimulus,

$$r(t_1, t_2) = \frac{Cov\left(rg_{\overrightarrow{PSV}_{encoding}(t_1)}, rg_{\overrightarrow{PSV}_{response}(t_2)}\right)}{\sigma_{rg_{\overrightarrow{PSV}_{encoding}(t_1)}} \sigma_{rg_{\overrightarrow{PSV}_{response}(t_2)}}} \quad , t_1 \in [0,2], t_2 \in [0, RT]\ sec$$

, with rg representing the ranking operator on the vector $\overrightarrow{PSV}$, and σ the variance of the vector. This produced a trial-specific two-dimensional similarity matrices, containing all the combinations of encoding (t_1) and response (t_2) time bins (Supplemental Figure 6d). The correlation coefficients r were then Fisher transformed, with the resulting coefficients following Gaussian distribution. The region-specific (amygdala and hippocampus) similarity matrices were averaged across trials within individual subjects, and used for group-level statistical analysis.

aSWR-locked memory reinstatement

Memory reinstatement during individual post-encoding time bins was computed by averaging the bin-specific similarity with the encoding period (200 time bins over 2 sec), resulting in a memory reinstatement time series. To obtain

the aSWR-locked memory reinstatement, we averaged the memory reinstatement within ± 250 msec around the individual aSWR peak times, separately for amygdala and hippocampus (Fig. 2.3a). We next tested whether the memory reinstatement is locked to aSWRs (Fig. 2.3b), by comparing the grand-average aSWR-locked reinstatement trace with an empirical null distribution obtained from Monte Carlo simulation. Specifically, we circularly randomly jittered the aSWR peak times within ± 500 msec window for 1000 times, obtaining an empirical null distribution of memory reinstatement strength.

To test whether the aSWR-locked reinstatement is associated with stimulus-induced arousal and later discrimination (Fig. 2.3c), we first derived the aSWR-triggered reinstatement, a metric taking the time-locked specificity relative to aSWR peak time into account. For every per-aSWR reinstatement trace around aSWR peak time, we circularly jittered the time as the procedure described above. This results in an empirical null distribution of reinstatement (i.e., correlation coefficient) for every time point around aSWR. We normalized the real reinstatement by z-scoring with mean and standard deviation of the null distribution. We referred to the resulting z-value as aSWR triggered reinstatement and it follows Gaussian distribution. We quantified the aSWR-locked reinstatement difference between the high/low arousal and between correct/incorrect discrimination at every time point by t-test, and corrected for the multiple comparisons using cluster-based nonparametric permutation test.

Specifically, we performed the group-level comparisons using paired t-test and identified contiguous time bins with the $p < 0.05$, defined as clusters. The t-values within each cluster were summed as the cluster statistics. We created an empirical null distribution by shuffling the conditional trial labels 1000 times where the maximum cluster statistics was identified for each permutation. It is considered as statistically significant if the real t-sum cluster statistics exceeded the 95% percentile of the null distribution.

Cross-structure joint aSWR-locked memory reinstatement

The cross-structure joint aSWR-locked memory reinstatement was obtained by calculating the outer product between the structure-specific reinstatement traces (hippocampus and amygdala) during post-encoding aSWR windows. The resulting joint reinstatement matrices were averaged across the individual aSWRs for each subject, separately for later correctly or incorrectly discriminated trials. To assess the statistical significance of joint cross-structure memory reinstatement, we performed a Monte Carlo simulation to generate an empirical null distribution by circularly jittering the aSWR peak times. The reinstatement significance was defined as exceeding the 95% percentile of null distribution (Fig. 2.3d).

Dual states analyses

Recorded periods were divided into low- and high-theta (3 - 10 Hz) or gamma (30 - 250 Hz) periods, based on the subject-specific power median split.

The aSWR occurrences are defined as the proportions of aSWRs occurring during each period. The aSWR occurrence comparisons between the low- and high-theta or gamma periods were performed using one-tailed Wilcoxon signed-rank test ($p < 0.05$; Supplemental Figure 9).

Mutual information

Mutual information (MI)(Cohen, 2014; Quian Quiroga & Panzeri, 2009) is a method for quantifying the amount of information shared between the variables of interest. In electrophysiology, MI is applied to test for the presence and directionality of information flow between the multiple time-series. We applied MI to assess the directional influence between the memory reinstatement in amygdala and hippocampus during the post-encoding aSWR windows (Fig. 2.3e). First, the structure-specific memory reinstatement traces from the amygdala and hippocampus were obtained around each aSWR event (± 250 msec; see aSWR-locked memory reinstatement). Next, we calculated the MI between the amygdala and hippocampal memory reinstatement traces, using the 200 msec bin size (10 msec step size), covering the ± 250 msec window around aSWR peaks. For each time bin, the reinstatement strength was binned into 10 bins (with uniform bin count), consistently across the subjects and conditions. The MI between the time series X and Y was defined as

$$MI(X;Y) = \sum_{i}^{n}\sum_{j}^{m} p(x_i, y_j)\log_2 p(x_i, y_j) - \sum_{i}^{n} p(x_i)\log_2 p(x_i) - \sum_{j}^{m} p(y_j)\log_2 p(y_j)$$

, where $p(x_i)$ and $p(y_j)$ represented the marginal probability of signals X and Y, $p(x_i, y_j)$ indicated their joint probability, while m and n represented the numbers of reinstatement strength bins for time series X and Y(Cohen, 2014; Quian Quiroga & Panzeri, 2009). To test the directionality of information flow, we calculated the time-lagged MI by shifting one time series relative to another across all the time bin combinations. The $MI_{AMY \to HPC}$ and $MI_{HPC \to AMY}$at individual time bins were defined as the mean of all the subsequent time-lagged MI bins in the other region(Cohen, 2014; Helfrich et al., 2019) . We defined the MI directional influence as the significant difference between the $MI_{AMY \to HPC}$ and $MI_{HPC \to AMY}$, assessed using Wilcoxon signed-rank test for each time bin. Correction for multiple comparisons was performed using the cluster-based nonparametric permutation test.

2.4 Results

2.4.1 Memory Discrimination is better for Emotional Stimuli

We performed simultaneous iEEG recordings from the amygdala ($n_{electrode}$ = 20) and the hippocampus ($n_{electrode}$ = 17, Fig. 2.2a) in 7 human subjects, while performing an emotional memory encoding and discrimination task(Leal et al., 2014; Zheng et al., 2019) (Methods, Fig. 1a). During the encoding stage, subjects were presented with a stimulus (image; stimulus encoding) and asked to rate the stimulus valence as negative, neutral, or positive (post-encoding/response). During the retrieval stage, subjects were presented with one of the 3 types of stimuli - Repeats (identical), Lure (slightly different) or Novel (stimuli not seen during encoding) - and classified each stimulus as "New" or "Old."

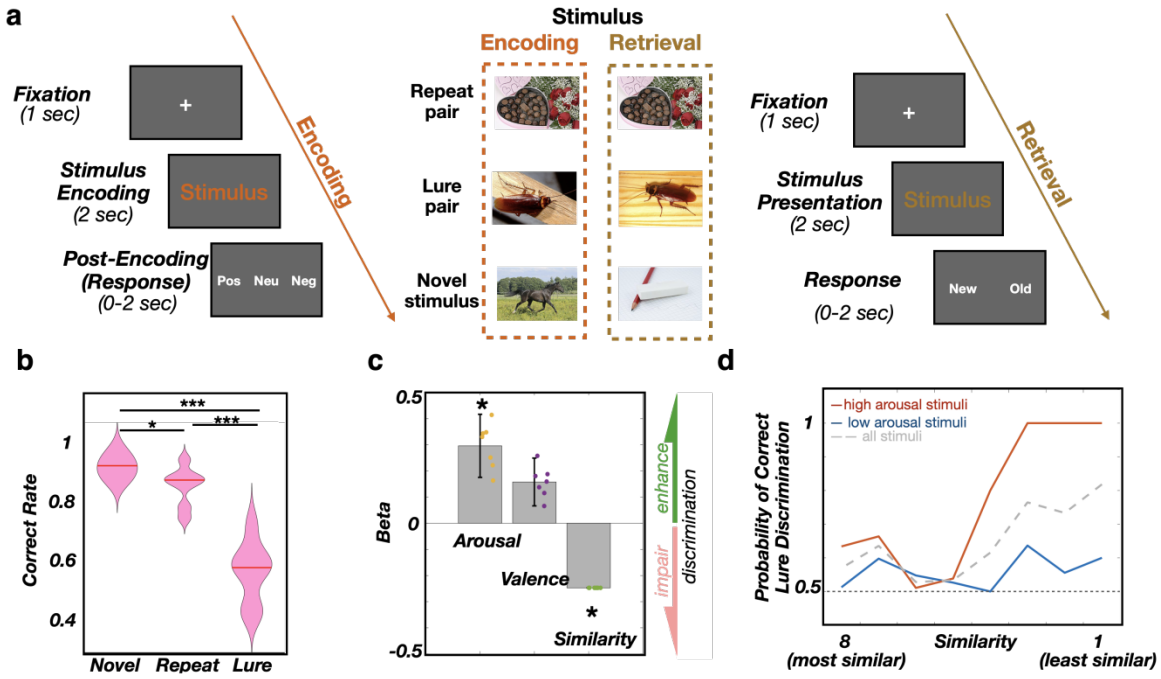


Fig. 2.1. Memory discrimination is more accurate for emotional stimuli.
a, Task structure: subjects are presented with an image (Stimulus encoding). Following presentation, they rate the valence of the image as negative, neutral, or positive (Post-Encoding/Response). Once all images are presented and rated, subjects are presented with 3 types of stimuli - Repeat (identical), Lure (slightly different) or Novel (stimuli not seen during encoding) - and classify each stimulus as "old" or "new." **b**, Correct discrimination is highest

for Novel stimuli (93.9 ± 1.4 %; median ± SEM), followed by Repeats (89.4 ± 2.4 %) and Lures (61.5 ± 3.7 %). Paired t-test: Novel vs. Repeat, *p = 0.016, t = 3.33, df = 6; Novel vs. Lure, ***p<0.001, t = 8.36, df = 6; Repeat vs. Lure, ***p < 0.001, t = 6.13, df = 6. **c**, Correct discrimination of Lure stimuli is positively associated with encoded stimulus-induced arousal (*p=0.047, β = 0.3 ± 0.12, t = 1.98, df = 452, logistic linear mixed-effect model) and valence (p = 0.137, β = 0.15 ± 0.09, t = 1.48, df = 452), while negatively associated with similarity (*p = 0.039, β = -0.24 ± 0.00, t = -2.06, df = 452). The β sign and magnitude indicate effect direction and strength, respectively. Dots correspond to individual subjects. **d**, Probability of Lure correct discrimination as a function of SI and stimulus-induced arousal. The solid line shows the actual proportion of 'New' responses (y-axis) as a function of Lure stimulus SI (x-axis) for low arousal (blue) or high arousal stimuli (red). The low/high arousal groups were created using the median split.

Memory discrimination is defined as the correct classification of: 1) Repeat stimuli as Old, 2) Novel stimuli as New, or 3) Lure stimuli as New. Subjects classified Repeat stimuli and Novel stimuli with high accuracy (Repeat: 89.4 ± 2.4%, Novel: 93.9 ± 1.4%; Fig. 1b). Memory discrimination accuracy was lower for Lure stimuli, relative to both Repeat or Novel stimuli (Lure: 61.5 ± 3.7 %; $p_{\text{Novel vs Lure}} < 0.001$, t = 8.36; $p_{\text{Repeat vs Lure}} < 0.001$, t = 6.13, paired t-test), reflecting similarity-induced memory interference. Indeed, there was a strong negative association between subjects' stimulus discrimination ability and stimulus similarity rating (p = 0.039, t = -2.06, see Methods, Fig. 1c-d). Stimulus-induced arousal (irrespective of valence) was associated with better memory discrimination, confirming previous reports(Cahill et al., 1998; Kensinger, 2009; Szőllősi & Racsmány, 2020) (p = 0.047, t = 1.98, Fig. 1c-d, Supplemental Fig. 1).

2.4.2 Post-encoding aSWR Occurrence Predicts Stimulus Emotional Content and Later Discrimination

We defined the post-encoding period as the interval between stimulus offset and subjects' stimulus valence rating response (Fig. 1a). We tested the association of post-encoding aSWR occurrence (i.e., the number of aSWRs) with the stimulus emotional content (stimulus-induced arousal and valence) and correct discrimination during retrieval. Higher post-encoding aSWR occurrence was associated with stimulus-induced arousal (p = 0.03, z = -2.2, Wilcoxon signed-

rank test, Fig. 2.2c) and also predicted correct discrimination during retrieval ($p = 0.03$, $z = -2.2$, Wilcoxon signed-rank test, Fig. 2.2c), but was not associated with stimulus valence ($p = 0.77$, $F(2, 15) = 0.25$, one-way ANOVA; Supplemental Fig. 3). Taken together, these results provide the first report of post-encoding aSWRs as a potential electrophysiological mechanism for enhanced memory discrimination of arousing stimuli, previously characterized at behavioral level(Kensinger, 2009; McGaugh, 2015; Szőllősi & Racsmány, 2020). Furthermore, the positive associations between aSWRs and stimulus-induced arousal/later discrimination were present in all individual subjects (Fig. 2.2c). The post-encoding response time (RT) did not differ based on stimulus-induced arousal ($p = 0.2$, $z = 0.7$, $RT_{high\text{-}arousal} = 0.8 \pm 0.1$ sec; $RT_{low\text{-}arousal} = 0.6 \pm 0.2$ sec) or later discrimination ($p = 0.25$, $z = 0.6$, $RT_{correct} = 0.7 \pm 0.2$ sec, $RT_{incorrect} = 0.7 \pm 0.3$, Wilcoxon signed-rank test). Therefore, the associations between stimulus-induced arousal or correct discrimination and post-encoding aSWR occurrence were unrelated to post-encoding duration. Associations between aSWR and stimulus-induced arousal/later correct discrimination accuracy were selective for the post-encoding time window. These relationships were absent for the stimulus encoding or the retrieval task stage ($p > 0.05$, Wilcoxon signed-rank test; Fig. 2.2c, Supplemental Fig. 3, 4). The aSWRs probability was significantly higher during low theta power periods (Supplemental Fig. 5), consistent with observations that cholinergic tone promotes theta oscillations and suppresses SWRs(Buzsáki, 2015; Jadhav et al., 2012). In addition, aSWRs did not overlap with increased broadband gamma power, suggesting that aSWRs are distinct from non-specific broadband power fluctuations(Bragin et al., 1999) (Supplemental Fig. 5).

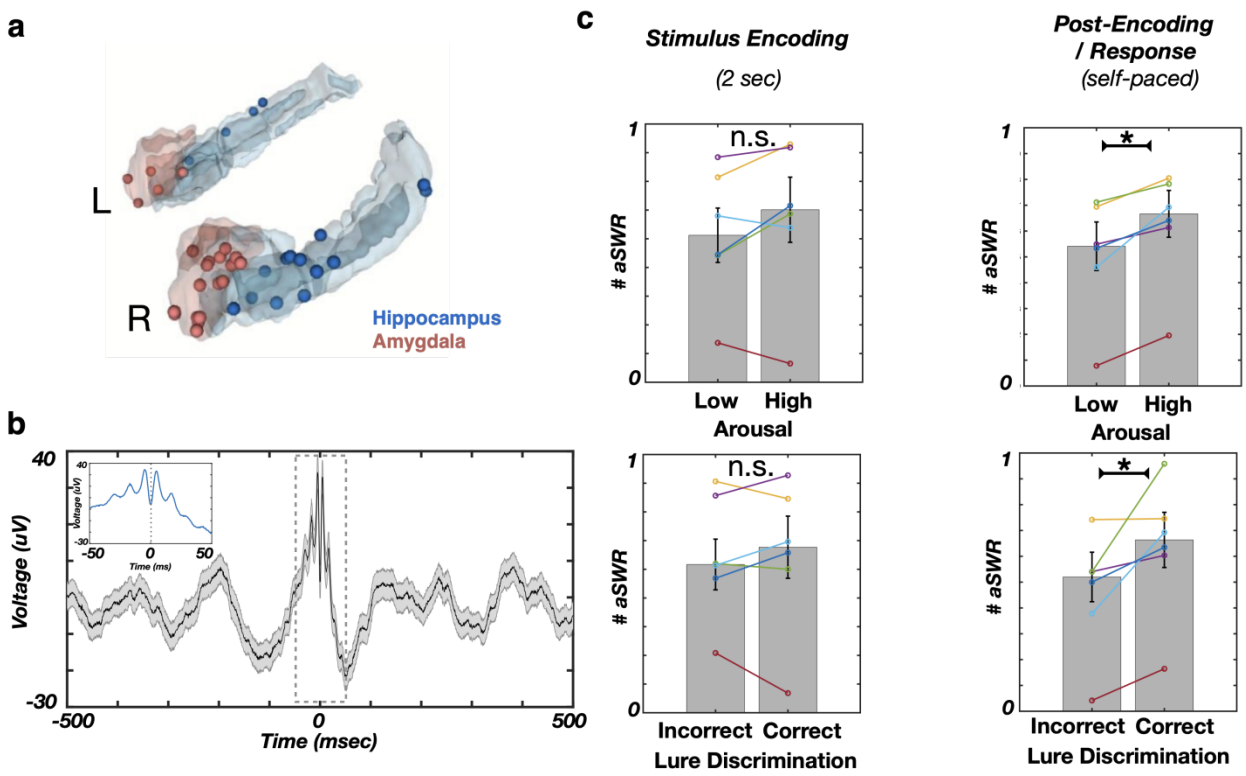


Fig. 2.2. The post-encoding aSWR occurrence predicts the stimulus-induced arousal and memory discrimination.

a, Reconstructed locations of hippocampal (blue) and amygdala electrodes (red). **b**, The aSWR grand average waveform (n = 4689 aSWRs in 6 hippocampal channels, 6 subjects). **c**, The aSWR occurrence is significantly higher following encoding of arousing (top right; *p = 0.03) and later correctly discriminated stimuli (bottom right, *p = 0.03). The aSWR occurrence was showing no conditional differences during stimulus encoding (left column, p's > 0.05).

2.4.3 Post-encoding Memory Reinstatement is Locked to aSWRs

Recent studies suggest that post-encoding memory reinstatement supports successful subsequent memory retrieval (Ben-Yakov et al., 2013; Sols et al., 2017). Meanwhile SWR is associated with reactivation of pre-established neuronal patterns(Genzel et al., 2020). We hypothesized that memory reinstatement during the post-encoding aSWR window could enhance later memory discrimination. Distinct neural populations have been proposed to represent individual stimuli, resulting in stimulus-specific high-frequency activity (HFA) patterns(Lopes-dos-Santos et al., 2018; Wixted et al., 2014). We, thus, quantified

memory reinstatement as the Spearman correlation between HFA power spectral vectors (PSVs), for each combination of the encoding-response time bins from the same trial (Supplemental Fig. 6). Next, we computed the average reinstatement activity during ± 250 msec around post-encoding aSWR peaks. The reinstatement significance was determined relative to a null distribution, obtained by circular jittering of aSWR timestamps. The post-encoding aSWR-locked memory reinstatement was stronger for arousing and correctly discriminated stimuli (Supplemental Fig. 7). To assess specific contributions of the amygdala and the hippocampus to this phenomenon, we calculated post-encoding memory reinstatement for each region, relative to aSWR peak (Fig. 2.3a). The significant reinstatement period in the amygdala consisted of two intervals, the first starting slightly earlier and overlapping with the hippocampal reinstatement (-105 to -50 msec), and a second period following the hippocampal reinstatement (40 to 200 msec). The significant reinstatement period in the hippocampus lasted from -100 to 50 msec (Fig. 2.3b). These results demonstrated region-specific timing of the post-encoding aSWR-locked memory reinstatement in the amygdala and the hippocampus. Next, we tested for the temporal compression(Genzel et al., 2020) of post-encoding aSWR-locked reinstatement (no compression, 2x, 4x, and 6x compression) and showed the strongest aSWR-locked reinstatement with no compression (Supplemental Fig. 8). We then analyzed the association of the post-encoding memory reinstatement with the stimulus-induced arousal and later discrimination. Remarkably, we observed a region-specific double dissociation. Specifically, the amygdala, not the hippocampus, showed a positive association between aSWR-locked memory reinstatements and the stimulus-induced arousal (AMY: -80 to -10 msec, $p = 0.035$; HPC: $p > 0.05$, see Methods; Fig. 2.3c). In contrast, the hippocampus, but not the amygdala, revealed a positive association between aSWR-locked memory reinstatement and later correct discrimination (AMY: $p > 0.05$; HPC: -15 to 90 msec, $p = 0.008$, see Methods; Fig. 2.3c). To summarize, post-encoding aSWR-locked memory reinstatements in the amygdala and the hippocampus followed distinct temporal dynamics and were

associated with reactivation of distinct aspects of encoded stimuli (i.e., the amygdala for stimulus-induced arousal and the hippocampus for later discrimination accuracy).

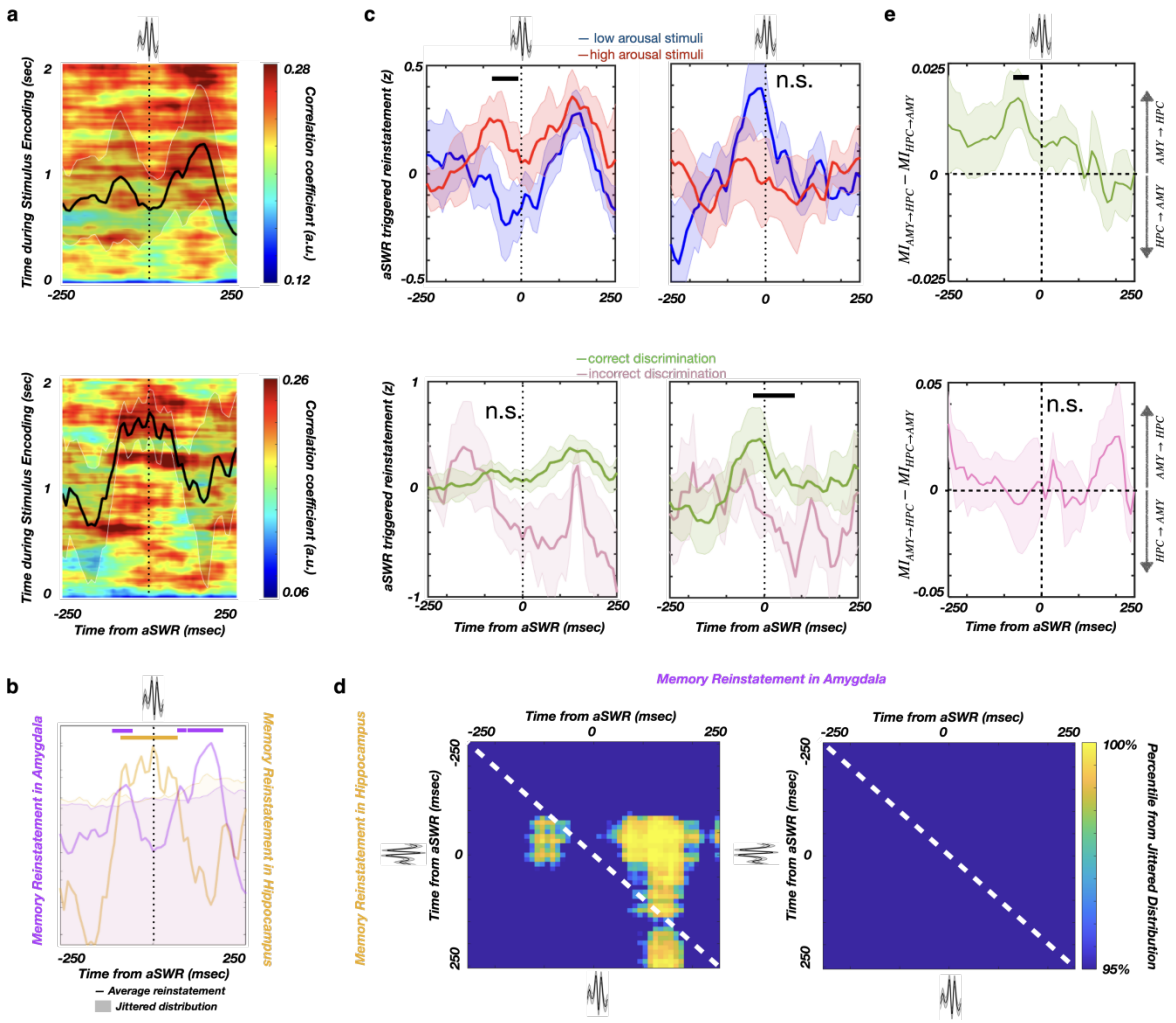


Fig. 2.3. Memory reinstatement in the hippocampus and amygdala around aSWR. a, aSWR-locked reinstatement in the amygdala (top) and hippocampus (bottom) during the post-encoding period (line and shaded areas represent the mean ± SEM). **b**, Reinstatement is greatest around the time of aSWRs as shown by comparison with the null-distribution (within ± 250 msec). Shaded areas denote the null-distribution 95% confidence interval. Reinstatement in the hippocampus overlaps with aSWR peak (orange), while reinstatement in the amygdala peaks prior to and after the aSWR (magenta). **c**, aSWR-locked reinstatement in the amygdala is stronger for arousing stimuli (top left, $p = 0.035$, see Methods) but is not associated with subsequent discrimination (bottom left, $p = 0.066$). Reinstatement in the hippocampus is robust for correctly discriminated stimuli (bottom right, $p = 0.008$, see Methods) but does not depend on stimulus-induced arousal (top right, $p > 0.1$). **d**, The aSWR-locked joint reinstatement in the hippocampus and amygdala for the correct (left) and incorrect (right) discrimination trials. Reinstatement in the amygdala starts 100 msec prior to the aSWR peak, followed by reinstatement in the hippocampus (-50 to 200 msec). There is no significant joint reinstatement

during incorrect discrimination trials, suggesting that the cross-structure joint reinstatement may be required for correct discrimination. **e,** Mutual information (MI) difference for the amygdala (AMY) and hippocampal (HPC) memory reinstatement time-courses, during the post-encoding aSWR windows (correct discrimination - top, incorrect discrimination - bottom). Positive values denote stronger AMY→HPC directionality. A temporal cluster of significant MI difference (AMY→HPC) is present before aSWR peak time(-70 to -30 msec) after encoding of correctly discriminated stimuli (top; $p = 0.038$, see Methods), indicating that hippocampal reinstatement is better predictable by amygdala reinstatement than vice versa. This effect is present only during the post-encoding period for correctly discriminated stimuli (top), but not for the incorrectly discriminated stimuli

2.4.4 Post-encoding aSWR-locked joint memory reinstatement in the hippocampus and amygdala is predictive of memory discrimination

In rodents, the coordinated memory reactivation in the amygdala and hippocampus during sleep SWRs is proposed to bind neuronal ensembles encoding emotional and spatial information, respectively(Girardeau et al., 2017). We reasoned that a similar interaction between the amygdala and the hippocampus exists in which cross-regional post-encoding aSWR-locked memory reinstatement facilitates later discrimination. We hypothesized that the reinstatement in both structures co-occurs during the same aSWR events and follows a consistent temporal dynamic. To test this, we separately computed aSWR-locked joint memory reinstatement for the correctly and incorrectly discriminated stimuli (Methods). A significant joint aSWR-locked memory reinstatement in the amygdala and hippocampus was present during the post-encoding period only for correctly discriminated stimuli (Fig. 2.3d; Supplemental Fig. 9). Specifically, the amygdala reinstatement preceded the hippocampal reinstatement by ~100 msec. Further, mutual information analysis showed a significant unidirectional influence from the amygdala to the hippocampus before aSWR peak (-70 to -30 msec, $p = 0.038$; see Methods; Fig. 2.3e). To conclude, aSWR-mediated coordination of memory reinstatement in the amygdala and the hippocampus promotes later successful discrimination.

2.5 Discussion

Rodent studies have implicated the SWRs in the retrieval and consolidation of emotional memory. However, it is unclear whether it supports the memory benefits of emotional experience(Trouche et al., 2020). Our study reveals an association of higher aSWR occurrence with stimulus-induced arousal and subsequent correct stimulus discrimination, providing direct evidence for aSWR-mediated strengthening of emotional memory. Interestingly, the higher aSWRs occurrence has been shown in rodents, after exposure to a novel or reward-associated context(Joo & Frank, 2018). Together, this suggests that aSWRs may play a general role in the selective enhancement of salient experiences(McGaugh, 2013).

Notably, such association is specific to the post-encoding period that starts immediately after memory encoding, when memory retrieval is essential to rate the emotional content of the stimuli. This finding supports theoretical assumptions that SWRs mediate both the retrieval of stored representation utilized in decision-making, and the strengthening of the same representation, contributing to memory consolidation(Joo & Frank, 2018).

Next, we aimed to discern the link between the aSWR-associated interaction between the amygdala and hippocampus during post-encoding and subsequent memory effect. We found the aSWRs were accompanied by memory reinstatement during the post-encoding period. Specifically, the reinstatement in the amygdala appears shortly before the aSWR peak and shows association with arousing stimuli, while the hippocampal reinstatement appears around the aSWR peak and shows associations with correct subsequent memory discrimination. Moreover, the co-occurrence of the amygdala and the hippocampal reinstatement during the same post-encoding aSWR events - with the amygdala reinstatement leading hippocampal by ~100 msec - is predictive of subsequent correct memory discrimination. This finding suggests that the

coordinated reinstatement in the amygdala and hippocampus during aSWR is responsible for combining emotional and contextual aspects of the memory(Girardeau et al., 2017; Trouche et al., 2020).

Both the joint-reinstatement and mutual information analyses further confirm the predictive validity of directional influence from the amygdala to the hippocampus before aSWRs on correct discrimination, establishing a link between the amygdala reinstatement and memory discrimination as a physiological mechanism of emotional memory enhancement. Together, our data support a model wherein the memory reinstatement in the amygdala, triggered by emotional stimuli, elicits amygdala-hippocampal aSWR-associated memory reinstatement, enabling the coordinated joint-reinstatement, which facilitates subsequent memory performance.

CHAPTER 3

Neural Dynamics of Pattern Separation in the Medial Temporal Lobe

3.1 Abstract

Episodic memory depends on pattern separation, the ability to discriminate between unique experiences. The neural dynamics of pattern separation was studied in the rodent hippocampus, with unclear behavioral relevance. Human imaging studies were consistent with behaviorally-relevant pattern separation, but the temporal resolution precluded testing the dynamical theories postulated by the rodent and theoretical work. We recorded the intracranial electroencephalogram (iEEG) from the human temporal lobe (amygdala, dentate gyrus/Cornu Ammonis3 (CA3), CA1 and parahipocampal cortex), during performance on mnemonic discrimination task. The stronger hippocampal representational similarity between the previously encoded and newly presented stimulus interferes with correct discrimination. The hippocampal representational dynamics is consistent with discrete attractors, characterized by abrupt transitions at sub-second time scale. Finally, higher representational dimensionality predicts correct discrimination, suggesting the code expansion as a mechanism implementing the pattern separation. This is the first

demonstration of behaviorally-relevant pattern separation dynamics at subsecond timescale in the human brain.

3.2 Introduction

The critical component of episodic memory is the ability to distinguish between unique, but similar experiences. For example, recognizing the nuanced difference between the two experiences allows us to separate the memories of different birthday parties, despite the shared structure and details. This ability is supported by the pattern separation - a neural computation that increases representational distance between the similar experiences (Marr, 1971; Treves, 1992; O'Reilly, 1994; McNaughton et al., 1987; Miller et al., 2019; Hainmueller and Bartos, 2020). In contrast, the failure of pattern separation is proposed as a pathophysiological mechanism underlying various cognitive and neuropsychiatric disorders (reviewed by Leal and Yassa, 2018).

The extensive theoretical and empirical work implies the hippocampal network in storage of unique memory representations (Marr, 1971; Treves, 1992; O'Reilly, 1994; McNaughton et al., 1987; Guzowski et al., 2004; Moser and Buzsaki, 2013). Within this network, the anatomical implementation of pattern separation was attributed to dentate gyrus (DG) and Cornu Ammonis 3 (CA3) hippocampal subfields (Berron et al., 2016; Bakker et al., 2008; Segel et al., 2012; Gilbert et al. 2001; McHugh, 2007; Wiskott et al., 2006; Sahay, 2019). The empirical support for DG/CA3 role in pattern separation is based on the spatial representation changes during exposure to environments with different degrees

of similarity, the phenomena known as remapping (Wills et al., 2005; Leutgeb et al., 2006; Leutgeb et al. 2007; Colgin et al, 2010; Jezek et al., 2011; Alme et al., 2015; Neunuebel et al., 2014; Knierim, 2016). Remapping could occur as a fast flickering between competing maps, suggesting the attractor-like organization of hippocampal spatial maps and possibly episodic memories in general (Hopfield, 1982; Morris and McNaughton, 1987; Redish, 2007; Fenton and Kelemen, 2010; Jezek et al., 2011; Kay et al., 2019). The mechanistic implementation of pattern separation underlying remapping was attributed to expansion recoding, an increase in the representation dimensionality during propagation through DG/CA3 network (Marr, 1969; Albus, 1971; Cayco-Gajic and Silver, 2019). It was hypothesized that remapping reflects the subjective perception of contextual change, but the behavioral relevance of remapping was difficult to establish in animal models (Fuhs and Touretzky, 2007; Colgin et al., 2008; Kubie et al., 2020; Sanders et al., 2020). Meanwhile, human fMRI studies have shown the evidence of behaviorally-relevant representational remapping between the trials in DG/CA3 (Julian and Doeller, 2021; Wanjia et al., 2021). However, establishing the relation of hippocampal remapping in humans with remapping dynamics observed in rodents and predicted by theoretical work requires the subsecond-level monitoring of remapping dynamics during individual trials.

To address this question, we recorded intracranial EEG (iEEG) from the human temporal lobe, while the subjects performed mnemonic discrimination task. In this task, the subjects were classifying image stimuli (Lures) as old or

new, based on the ability to discriminate them from highly similar, previously presented stimuli (Targets). We hypothesized that the pattern separation neural dynamics in DG/CA3 could be identified by contrasting the neural activity during the correct and incorrect trials (correct rejections and false alarms, respectively). The representational similarity between the Target and Lure trial pairs in DG/CA3 was significantly higher for false alarm, relative to correct rejection trials, suggesting that the balance between pattern separation and completion determines the perceptual similarity. Neural trajectory through representational space during discrimination trials was flickering between the states of high and low similarity to Target stimulus. The states were separated by a large representational distance and the individual state visitation was predictive of stimulus discrimination behavioral outcome. Finally, the representational dimensionality during correct rejection trials was higher relative to false alarm trials, implying the expansion recoding as a mechanism underlying successful pattern separation (Cayco-Gajic and Silver, 2019). Overall, these findings represent the first demonstration of behaviorally-relevant hippocampal remapping at subsecond timescale in the human brain, consistent with a long-standing theoretical prediction of expansion recoding as an underlying mechanism.

3.3 Materials and Methods

Subjects

Data were obtained from 8 subjects (3 females) with pharmacoresistant epilepsy (age 38 ± 18, mean ± SD; see Table 1 for individual subject demographics), during pre-surgical evaluation at the University of California Irvine, Medical Center (UCIMC). The localization of implanted electrodes was determined strictly based on the individual subject clinical situation. All the procedures were approved by the UCIMC Institutional Review Board and each subject provided informed consent prior to data collection.

Mnemonic discrimination task

The mnemonic discrimination task (Fig. 3.1A) consisted of the encoding (n = 148) and discrimination trial blocks (n = 290). Encoding trials consisted of a fixation epoch, when a fixation stimulus (white cross) was presented for 1000 ms, followed by the presentation of image stimulus at the center of the screen (2000 ms) and the response epoch (self-paced, up to 2000 ms). During response epoch, subjects rated the emotional valence of the presented stimulus as either positive, negative or neutral, using button press. Discrimination trials followed the same temporal structure - fixation epoch (1000 ms), stimulus presentation (2000 ms), and the response epoch (self-paced, up to 2000 ms). During the response epoch, subjects were asked to indicate (using button press) whether the presented stimulus is identical to one of the stimuli from the encoding block (Old) or not (New). The subjects were instructed that a similar but not identical stimuli should also been responded as a New stimuli.

Three types of stimuli were presented during the discrimination block (Fig. 1A): Repeat - the stimulus identical to one of the stimuli from encoding trials (n = 54), Lure - a stimulus similar to the one previously seen, but not identical (n = 97) and Novel - a stimulus with no similarity to any of the previously presented stimuli (n = 139).

Statistics

For the group level analyses, the linear mixed-effect model (LME) was implemented, in the form of $y = \beta X + uZ + \varepsilon$, with y denoting the response on individual trial , $X = [x_1, x_2, \dots, x_K]^T$ denoting K task variables of interest, $Z = [z_1, z_2, \dots, z_L]^T$ denoting L group variables, β denoting the fixed-effect coefficients, u denoting the random-effect coefficients, and ε denoting the random noise. Subject identity was the grouping variable.

Behavioral Analyses

The correct rate were first calculated as the portion of correctly responded trials over all trials per condition (P(Old | Repeat), P(New | Lure), P(New | Novel)). We then compared the correct rate difference between task conditions using paired t-test (Fig. 3.1C).

Data collection

The behavioral task was administered using PsychoPy2 (Version 1.82.01) software[2] installed on an Apple MacBook Pro. The task laptop was placed in front of the patient at a comfortable distance. The intracranial EEG (iEEG) signal was digitized and recorded using a Nihon Kohden acquisition system (256 channel amplifier, model JE120A), with an analog highpass filter (0.01 Hz cut-off) and sampling frequency of 5000 Hz.

Electrode localization

Pre-implantation structural T1-weighted MRI (pre-MRI) and post-implantation MRI (post-MRI) or CT (post-CT) scans were used for electrode localization. The pre-MRI and post-MRI (or post-CT) were co-registered with translations and rotations in x, y and z directions, using Advanced Normalization Tools (ANTs; https://stnava.github.io/ANTs/). A high-resolution anatomical template with the labels of medial temporal lobe subfields was used to guide the localization of individual electrodes. The template was resampled with 1 mm isotropic and aligned to pre-MRI using ANTs Symmetric Normalization, producing a subject-specific template. The electrode locations were identified by comparing the subject-specific template subfield areas with electrode artifacts. The localization results were reviewed by the neurologist (J.J.L.). For the visual illustration, we collected the electrode coordinates in the MNI space across all the subjects and rendered on the Colin27 template brain (Lever et al. 2009).

Preprocessing

All the analysis was performed using the custom-written MATLAB code (Version 9.7) or FieldTrip Toolbox [3]. The digitized EEG signal was down-sampled to 2000 Hz, demeaned and high-passed filtered (0.3 Hz). The power spectrum density (PSD) of the signal was computed using the multitaper method with the Hanning window (ft_freqanalysis.m function in FieldTrip). The 60 Hz line noise and its harmonics were removed using a FIR bandpass filter (ft_preprocessing.m function in FieldTrip). The signal from each channel was re-referenced by subtracting the signal from the nearest white matter channel on the same depth electrode. The epileptiform discharges were manually marked by an epileptologist (J.J.L.; ft_databrowser.m function in FieldTrip). The channels with severe line noise contamination and trials containing epileptiform discharges were excluded from further analyses.

Unsupervised decomposition of intracranial EEG

Ensemble empirical mode decomposition (EEMD) was applied to decompose the iEEG signal into distinct modes (EEMD MATLAB package; https://github.com/leeneil/eemd-matlab.git). The EEMD procedure consists of breaking down a non-stationary signal into elementary components, referred to as intrinsic mode functions (IMFs; Fig. S1; Oostenveld et al., 2011). The EEMD algorithm is iteratively applied with n=100, while adding white noise to prevent mode mixing (Oostenveld et al., 2011; Ray and Maunsell 2011). Such

decomposition solely depends on the signal's intrinsic properties, without prior assumptions on the mode structure (Oostenveld et al., 2011; Ray and Maunsell 2011; Lopes-dos-Santos et al., 2018). The resulting IMFs consistently capture several canonical spectral features across all recorded anatomical structures, with center frequencies in delta (IMF1, 1.5 Hz), theta (IMF2, 2.5 Hz), alpha (IMF3, 8 Hz), beta (IMF4, 20 Hz), gamma (IMF5, 57.5 Hz) and high-gamma range (IMF6, 159.5 Hz), as well as noise term (IMF7) with frequency > 300 Hz, and ultra-slow component with center frequencies < 0.5 Hz (Fig. S1). The gamma range activity was then reconstructed by summing the IMFs with center frequencies in the 30 - 300 Hz range. This procedure avoids the contamination of gamma activity by low-frequency harmonic artifacts structure (Oostenveld et al., 2011; Lopes-dos-Santos et al., 2018).

Time-frequency representation

The instantaneous spectral power at each time bin was derived from the reconstructed gamma time series using 'Superlet' transformation, which is a wavelet-based approach (Bârzan, 2020; Moca et al., 2021), resulting in a trial-specific time-frequency matrix (Fig. S1). Specifically, the reconstructed iEEG time series was convolved with a set of Morlet wavelets,

$$P_{f,n}(t) = \left|\psi_{f,n} * x(t)\right|, n = 2,3, \dots ,10$$

$$\psi_{f,n} = \frac{1}{B_n\sqrt{2\pi}} e^{-\frac{t^2}{2B_n^2}} e^{j2\pi ft}, where\ B_n = \frac{n}{5f}$$

parametrized by a range of cycle numbers (n = 2, 3,…, 10) at a given frequency *f*, and the geometric mean was computed for individual time-frequency bins.

$$\hat{P}(f,t) = \sqrt[9]{\prod_{n=2}^{10} P_{f,n}(t)}$$

This approach facilitates the detection of transient oscillations in narrow frequency bands (Bârzan, 2020; Moca et al., 2021).The wavelet center frequencies used were in a range 30 - 280 Hz, with 1 Hz step size. The cycle number range of 2-10 was based on the previous literature[9]. To avoid the edge effects, the procedure was applied on the entire recording, followed by segmentation of the resulting matrix into trial epochs (starting 1000 ms prior and ending 2000 ms after the stimulus presentation onset). Finally, the power in each individual frequency was normalized at the individual trial level by z-transformation and correction for average pre-trial baseline power (-1000 - 0 ms, relative to stimulus onset)[9].

Representational Similarity Analysis

To quantify the representational similarity (Lohnas et al., 2018; Norman et al., 2019; Yaffe et al., 2014; Zhang et al., 2018) between the stimulus encoding and discrimination trials, the instantaneous spectral power was first binned into a 200 ms time windows, with 40 ms (80% overlap). The Spearman's correlation was

computed as a measure of similarity between the two power spectral vectors (PSVs) for encoding and discrimination trials I and J, respectively

$$\overrightarrow{PSV}_{encoding,I}(t_1) = \left[z_1(t_1), \dots, z_{n_f}(t_1)\right]_{encoding,I}$$

$$\overrightarrow{PSV}_{discrimination,J}(t_2) = \left[z_1(t_2), \dots, z_{n_f}(t_2)\right]_{discrimination,J}$$

$$r_{I,J}(t_1, t_2) = \frac{Cov\left(rg_{\overrightarrow{PSV}_{encoding,I}(t_1)}, rg_{\overrightarrow{PSV}_{discrimination,J}(t_2)}\right)}{\sigma_{rg_{\overrightarrow{PSV}_{encoding,I}(t_1)}} \sigma_{rg_{\overrightarrow{PSV}_{discrimination,J}(t_2)}}} \quad , t_1, t_2 \in [0,2]\ sec$$

, with rg denoting the vector ranking operator, and σ denoting the vector variance. This procedure generated a two-dimensional representational similarity map for each stimuli pair, spanning the entire encoding and discrimination trial (Fig. S2). Next, the Spearman correlation coefficients were z-transformed using Fisher transformation.

The stimulus-specificity of neural representations on Repeat trials (Fig. 3.2A; Fig. S2) was tested by comparing the representational similarities of the repeated presentations of the same stimuli (Same), with the randomly paired stimuli (Different), using one-tailed paired t-test (Fig. 3.2A; Fig. S2). The representational similarity matrices at individual electrode level were averaged over conditions, followed by averaging at the subject level and statistical testing at the group level (one-tail t-test; $p < 0.05$). Correction for multiple comparisons was done using a non-parametric cluster-based permutation (Maris and Oostenveld, 2007), with

random shuffling of stimulus identity (n = 1000 permutations). The analyses were performed separately for each ROI (amygdala, hippocampus, peri-hippocampal structure).

The representational similarity of the Lure encoding-discrimination trial pairs (Fig. 1A) was compared using the same procedure, except that the comparisons were made between the presentation of highly similar Target-Lure pairs (Similar) and randomly paired Target-Lure pairs (Different).

State-space analysis

State-space analysis was performed on encoding-discrimination Lure trial pairs (Fig. 1A). We designed an algorithm to predict the neural representation at each time point during the discrimination trial, by investigating its corresponding stimulus-specific memory representation during encoding trial. The window of unique memory representation (encoding window) was defined based on the peak times of stimulus-specific representational similarity across the Repeat trial pairs from all subjects (0.95 ± 0.51 sec; median ± SD; Fig. S3). The rationale for defining the encoding window based on the pooled trials was to use the fixed window width across subjects. Next, the PSVs from the encoding window (1000 ms), pre-discrimination trial baseline (1000 ms; cross-fixation) and discrimination trial (2000 ms) were concatenated for each encoding-discrimination trial pair (Fig. S4B). The pre-discrimination trial baseline was used to control for the possibility that representational distance between the encoding

and discrimination trials is attributable to change in temporal context (Howard and Kahana, 2002; Polyn et al., 2009; Folkerts et al., 2018). Next, the principal component analysis (PCA) was performed on concatenated TFR and the resulting vector was projected into a low-dimensional space. Consistent with previous studies, the dimensionality of the principal component space is determined by the equation

$$D = \frac{\left(\sum_{i=1}^{N_f} \lambda_i\right)^2}{\sum_{i=1}^{N_f} \lambda_i^2}$$

, where λ_i is the i-th eigenvalue from singular value decomposition (SVD). Theoretically, the resulting dimensionality covers the large proportion of variance (>80%) in the high dimensional space. So that each data point were projected into a stimulus-specific D-dimensional subspace (Fig. S4C).

A k-nearest-neighbor (k-NN) algorithm was used to classify each discrimination time point into one of the two sub-states (Encoding, Baseline) using data point during encoding trial and pre-discrimination baseline trial as training samples. Next, the states that were outside of the assigned state boundary (as defined by the 95th percentile) were classified as Outside states (Fig. S4C). Specifically, we calculated an within-category pairwise distance distribution for each encoding data point. The 95th percentile of the distance distribution was then identified as a radius to draw the boundary corresponding to one encoding data point (Fig. S4D). The combination of the boundaries defined by each encoding data point

constructed an enclosed area surrounding the encoding data points. The same procedure was used to draw the boundary defined by pre-discrimination baseline data points (Fig. S4C). As a result, the designed algorithm identified each discrimination time point into three possible sub-states (Encoding, Baseline and Outside).

The state-space velocity was defined as the Euclidean distance between the PC vectors at neighboring time-points. The state-space transitions are defined as the timepoints of transition between any of the sub-states (T_0; Fig. 3.4D). The transition-locked state-space velocity was calculated by averaging the velocity during the time-windows centered at T_0. To assess the velocity changes at state transitions, the average velocity locked to the state transitions (encoding to outside and opposite direction +/- 250 ms). To assess the statistical significance of higher velocity at state transition points, we performed a Monte Carlo simulation to generate an empirical null distribution by circularly shifting transition time points with n=1000 times. The velocity significance was defined as exceeding the 95% percentile of null distribution.

Transition probability analysis

To assess the sub-state transition dynamic between sub-states (Encoding, Baseline and Outside), the transition probability was computed at individual trial level, for each sub-state combination and direction. P was defined as the instance of unidictional transition between two states over all possible transition

time points duing one trial . Next, the transition probabilities were compared between the trials with different behavioral outcomes (CR and FA), using the linear mixed effect model (Fig. 3.4E). Correction for multiple comparisons was performed using Bonferroni method, based on the number of state combinations.

Expansion recoding analysis

Expansion recoding is defined as the increase in signal dimensionality during input propagation through a neural network. Signal dimensionality was computed by performing the PCA on the TFR (Fig. 3.5), followed by singular value decomposition (SVD) and evaluating the dimensionality using the following equation.

$$D = \frac{\left(\sum_{i=1}^{N_f} \lambda_i\right)^2}{\sum_{i=1}^{N_f} \lambda_i^2}$$

Signal dimensionality was computed starting from the time window including the Encoding and Baseline periods (2 sec), increasing the window size at 50 ms increments, until covering the entire Encoding, Baseline and Discrimination windows (4 sec). Dimensionality during discrimination trials was compared between the conditions (CR and LA; Fig. 3.5B). Dimensionality expansion was defined as the first derivative of dimensionality trace (Fig. 3.5B). Dimensionality expansion was compared for the discrimination trial time windows before and

after the center of significant representational similarity cluster (Fig. 3.5C), separately for hippocampal subfields (DG/CA3 and CA1) and behavioral outcomes (CR and LA).

Non-parametric cluster-based permutation

When applying a statistical test multiple times, it may falsely provide positive results, and any of its false positives may result in a false conclusion of the overall hypothesis. To address this issue, we applied a cluster-based permutation test (Maris and Oostenveld, 2007). For the RSA, stimulus identity were randomly permuted 1000 times (Same Fig. 3.2 and Fig. 3.3). For each permutation, the representational similarity was computed and contiguous encoding-discrimination temporal windows with $P < 0.05$ were identified as individual clusters. Within each cluster, the t-value was summed (cluster-specific t-sum) and the maximum t-sum was obtained for permutation. This resulted in an empirical null distribution, and the true cluster statistics were compared against the 95% percentile of the null distribution for testing the statistical significance.

3.4 Results

3.4.1 Task Paradigm and Behavior Result

To investigate the neural dynamics underlying pattern separation, we employed a emotional memory encoding and discrimination task (EMOP; Fig. 3.1a, Leal et al., 2014; Zheng et al., 2019), while recording intracranial

electroencephalogram (iEEG; Fig. 3.1b) in the amygdala (n = 23), hippocampus (n = 21), and peri-hippocampal areas (n = 19; including parahippocampal, perirhinal, and entorhinal cortex) in 8 pre-surgical patients with epilepsy (3 females, see Table S1 for detailed demographics). During encoding trials, stimuli were presented on a computer screen for 2 s. Next, the subjects were instructed to classify the stimulus valence as negative, neutral, or positive. During discrimination trials, subjects were presented with three types of stimuli, classified based on the relation to encoded stimuli: 1) Repeats - stimuli identical to one of the encoded stimuli, 2) Lures - stimuli highly similar, but not identical to one of the encoded stimuli and 3) Novel - an entirely new stimulus, not similar to any of the encoded stimuli. Subjects were instructed to classify the stimulus as 'Old' or 'New', depending if the stimulus was perceived as one of the stimuli presented during encoding or not. Subjects performed the task with a high accuracy, classifying Novel stimuli as New (94%, $P(\text{New} \mid \text{Novel}) = 0.94 \pm 0.01$; Fig. 3.1c) and Repeat stimuli as Old (90%, $P(\text{Old} \mid \text{Repeat}) = 0.90 \pm 0.02$). As expected, the discrimination accuracy for Lure stimuli was lower than for Novel or Repeat stimuli (61%, $P(\text{New} \mid \text{Lure}) = 0.61 \pm 0.03$; $p_{\text{Novel vs Lure}} < 0.001$, $t = 8.36$;

$p_{\text{Repeat vs Lure}} < 0.001$, $t = 6.13$, paired t-test), suggesting memory interference between similar stimuli.

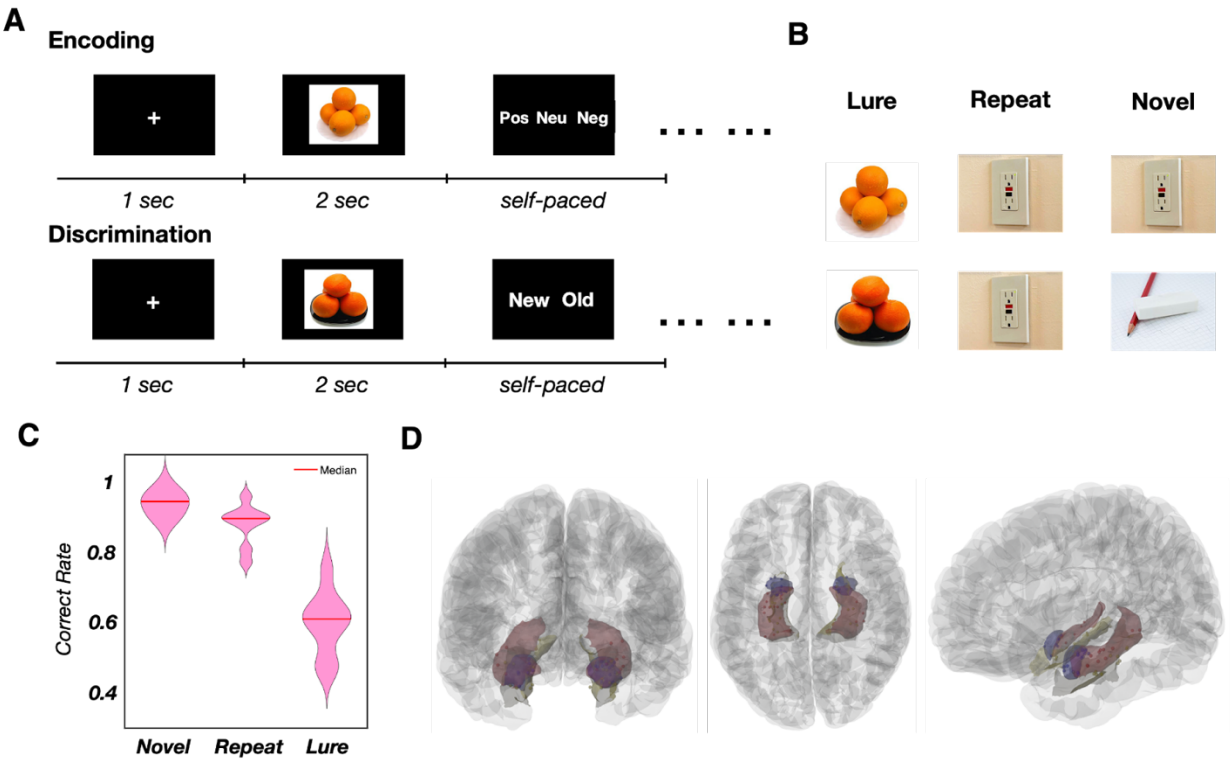


Fig. 3.1. Behavioral task and electrode localization.
a, Schematic task design. During the Encoding trials, the subjects were asked to rate the valence of the presented stimuli (Targets) as negative, neutral, or positive. **b,** During the discrimination phase, three types of stimuli were presented: Repeat - a stimulus identical to one of the previously presented Target stimuli; Lure - a stimulus partially overlapping with the one of Target stimuli; and Novel - a stimulus without similarity to any of the Target stimuli. The subjects were instructed to discriminate whether the presented stimulus (Repeat, Lure or Novel) was identical as one of the Target stimuli (Old) or different from any of the Target stimuli (New). **c,** Discrimination accuracy (%) for Novel, Repeat, and Lure stimuli (N = 8 subjects, median ± SEM). Discrimination accuracy was high for Novel (94.5 ± 1.3%, Novel recognized as New) and Repeat stimuli (89.6 ± 2.1%, Repeat recognized as Old). The performance was lower for Lure stimuli (61.0 ± 3.3%, Lure recognized as New), reflecting the memory interference. **d,** Intracranial depth electrode locations across all the subjects. The iEEG signal was simultaneously recorded in three ROIs: hippocampus (blue, n=21), amygdala (red, n=23), and peri-hippocampal regions (yellow, n=19), including the entorhinal cortex (EC), parahippocampal cortex (PHC) and perirhinal cortex (PRC).

3.4.2 Stimulus-specific memory representation in the hippocampus

The iEEG signal was parsed into distinct frequency bands (intrinsic mode functions; IMFs), using ensemble empirical mode decomposition (EEMD; Wu and Huang, 2009; Lopes-dos-Santos et al., 2018; Fig. S1, see APPENDIX B). The EEMD is a data-driven unsupervised approach, particularly suited for segregating signals from nonlinear and non-stationary time series, such as the iEEG (see Methods). The IMFs with center frequencies within gamma range (30 - 300 Hz) were extracted following the EEMD, due to the high correlation of gamma power with spiking activity (Ray and Maunsell, 2011) and hippocampal memory representations (Staresina et al., 2016.; Pacheco et al., 2019). The gamma IMF time series were summed and the resulting signal was spectrally decomposed using 'Superlet' transform, resulting in individual trial time-frequency matrices (see Unsupervised decomposition of intracranial EEG, Methods). Next, the representational similarity analysis (RSA, Kriegeskorte et al., 2008; McKenzie et al., 2014; Nili et al., 2014) was applied to quantify the similarity between the neural activity patterns at encoding and discrimination trials for identical stimuli (Repeats; e.g. A_{enc} and A_{ret}). Representational similarity was defined as the Spearman correlation of spectral power vectors for all encoding and discrimination trial time bin combinations, producing a unique representational similarity matrix for each encoding-discrimination stimulus pair (Fig. S2, see APPENDIX B).

The necessary prerequisite for studying the dynamics of pattern separation is the ability to identify unique stimulus representations in the hippocampus. We defined the unique stimulus representation as the higher encoding-discrimination representational similarity for the repetitions of the same stimuli (Same) than for encoding-discrimination trial pairs of randomly paired different stimuli (Diff). The representational similarity was quantified by correlating the neural activity during encoding (Enc) and discrimination (Ret) trials, for identical stimuli (RSA_{Same}; e.g. stimuli A_{Enc} and $A_{Ret,}$) or different stimuli (RSA_{Diff}; e.g. A_{Enc} and B_{Ret}). Indeed, the RSA revealed a stimulus-specific representation in the hippocampus ($RSA_{Same} > RSA_{Diff}$; $p = 0.028$, nonparametric cluster-based permutation test, 1000 permutations, Fig. 3.2a). In contrast, there was no stimulus-specific representation in the amygdala and peri-hippocampus, reflected by no significant difference between the RSA_{Same} and RSA_{Diff} in those regions ($p > 0.05$, nonparametric cluster-based permutation test, 1000 permutations; Fig. 3.2a). To assess if the RSA difference in hippocampus is driven by increased similarity of the Repeat pairs or decreased similarity of Different pairs, we compared the representational similarity of these trial groups during the significant temporal similarity cluster (Fig. 3.2a, enclosed by black line). The average hippocampal representational similarity within the significant temporal window in the hippocampus was higher for the Same trial pairs than for Different trial pairs ($p < 0.005$, one-tailed paired t-test; Fig. 3.2b). There was no significant difference in either amygdala or parahippocampus ($p > 0.05$, one-tailed paired t-test; Fig. 3.2b). Overall, these results confirm the previous

theoretical and empirical notions of unique memory representations in the hippocampus (Marr, 1971; Treves, 1992; O'Reilly, 1994; McNaughton et al., 1987; Guzowski et al., 2004; Moser and Buzsaki, 2013; Lohnas et al., 2018; Pacheco-Esteban et al., 2019; Popham et al., 2021; Gelbard-Sagiv et al., 2008) and validate the approach for identifying those representations.

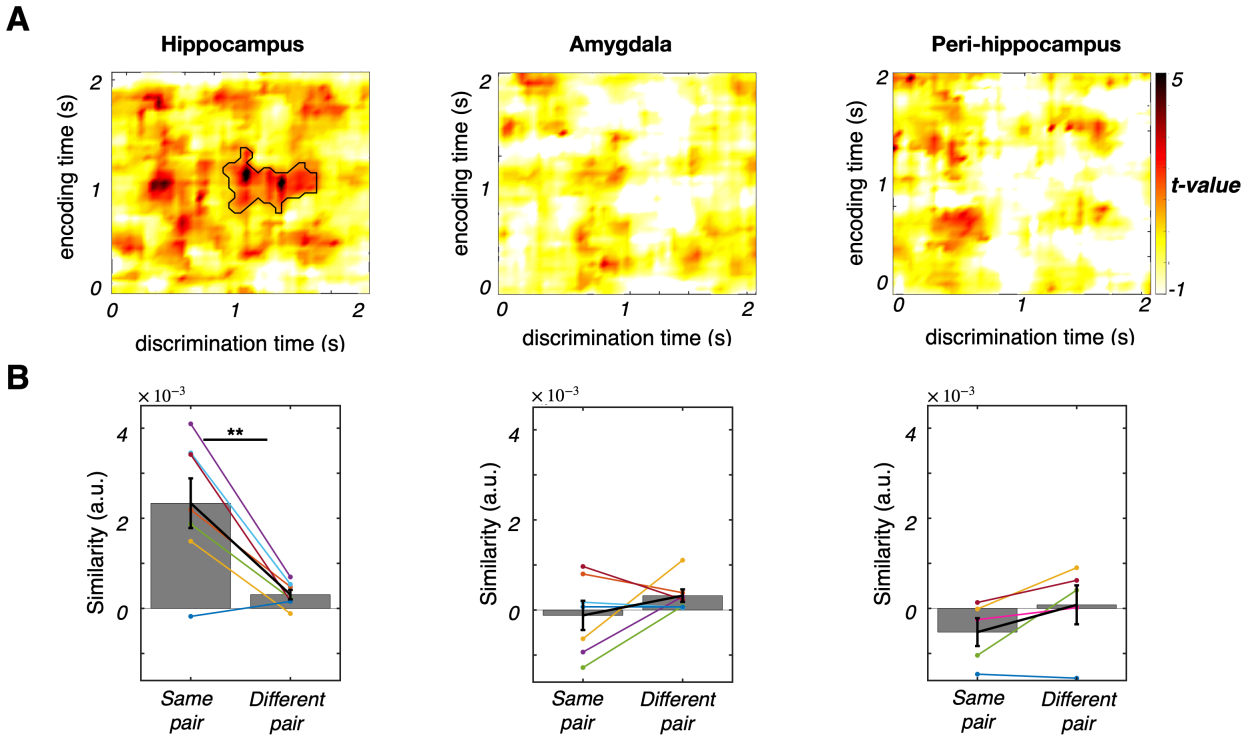


Fig. 3.2. Stimulus-specific neural representation in the hippocampus.
a, Regional t-maps showing the similarity between the multiple presentations of the same stimuli (Repeats). The temporal window of significant stimulus-specific neural representation in the hippocampus (left) is encircled by black line (p=0.028, nonparametric cluster-based permutation test, Same > Different). There was no significant stimulus-specific representation in the amygdala (middle) or peri-hippocampus (right), suggesting the hippocampus as the location of unique memory representations. **b,** The average representational similarity during a significant temporal window shown in A is significantly higher for the Same-pair trials, relative to Different-pair trials (p=0.003, t(6)=4.126, one side paired t-test) and the difference is consistent across the individual subjects (color coded). There was no significant difference during the same temporal window in the amygdala (middle, p=0.85, t(6)=-1.17, one side paired t-test) or peri-hippocampus (right, p=0.9574, t(6)=-2.27, one side paired t-test).

3.4.3 Stimulus-specific neural representation in the hippocampus predicts successful or failed discrimination

The peak stimulus-specific representational similarity was highest between the 0.95 ± 0.51 (mean +/- SD) sec during encoding and 1.05 ± 0.50 sec during discrimination (Fig. S3, see APPENDIX B). Next, we hypothesized that during Lure trials (Fig. 3.1b), the degree of representational similarity in the hippocampus will predict the behavioral outcome, i.e. whether the Lure item is correctly classified as New or misclassified as Old. Based on this prediction, the stimuluss representation during the failed discrimination trials (FA) would show higher similarity to respective encoding trials, compared to stimulus representation on correct discrimination trials (CR). This would reflect the reinstatement of the originally encoded stimulus, triggered by the presentation of the Lure stimulus on the FA trials. In contrast, lower representational similarity on CR trials would reflect decorrelation of Lure stimulus sensory input from the original stimulus representation. These two dynamical patterns are consistent with the pattern completion and pattern separation mechanisms, respectively. The dentate gyrus (DG) has been implicated in transformation of similar cortical inputs into distinct memory representations. Therefore we predicted that neural activity patterns in DG/CA3, but not in CA1 hippocampal subfield, will show decorrelated patterns for correctly discriminated (CR), compared to incorrectly discriminated (FA) Lure stimuli.

To test these hypotheses, we first quantified representational similarity between the similar and different stimulus pairs in Lure trials. When subjects misclassified Lure images as Old (FA), the DG/CA3 representational similarity was significantly higher for the pairs of similar, relative to unrelated stimuli ($RSA_{Similar} > RSA_{Different}$; $p = 0.042$; nonparametric cluster-based permutation test, 1000 permutations). The temporal cluster of significant difference spans 0.7 - 1.3 s following the encoding onset and 0.8 - 1.1 s following the discrimination onset (Fig. 3.3a). This result suggests that the failure of mnemonic discrimination of highly similar stimuli is associated with significant representational similarity. In contrast, there was no significant representational similarity between the similar stimuli that were correctly perceived as different (CR condition; $p > 0.05$; nonparametric cluster-based permutation test, 1000 permutations). This dissociation was selective for the DG/CA3, as there was no significant representational similarity between the similar items in the amygdala or parahippocampus, regardless of the correct stimulus discrimination ($p > 0.05$; nonparametric cluster-based permutation test; Fig. 3.3b, S1). Thus, the presence of correlated neural representation in DG/CA3 predicts the false recognition of the Lure item as identical (FA). Additionally, the timing of this representational similarity during the encoding is strikingly similar to the timing of stimulus-specific representation in repeated stimuli, suggesting that during lure false alarm, the DG/CA3 network erroneously reinstates the representation of a similar but non-overlapping stimulus, resulting in the lack of discrimination between those stimuli (Fig. S2). The average correlation coefficient of same-pair

and different-pair within tROI during FA and CR in the hippocampus across all the subjects. The same-pair correlation was significantly higher than different-pair during FA. The same pair correlation during CR was significantly lower than during FA. Moreover, the neural representation during FA is more similar than during CR with its encoded stimulus ($p = 0.023$, paired t-test; Fig. 3C).

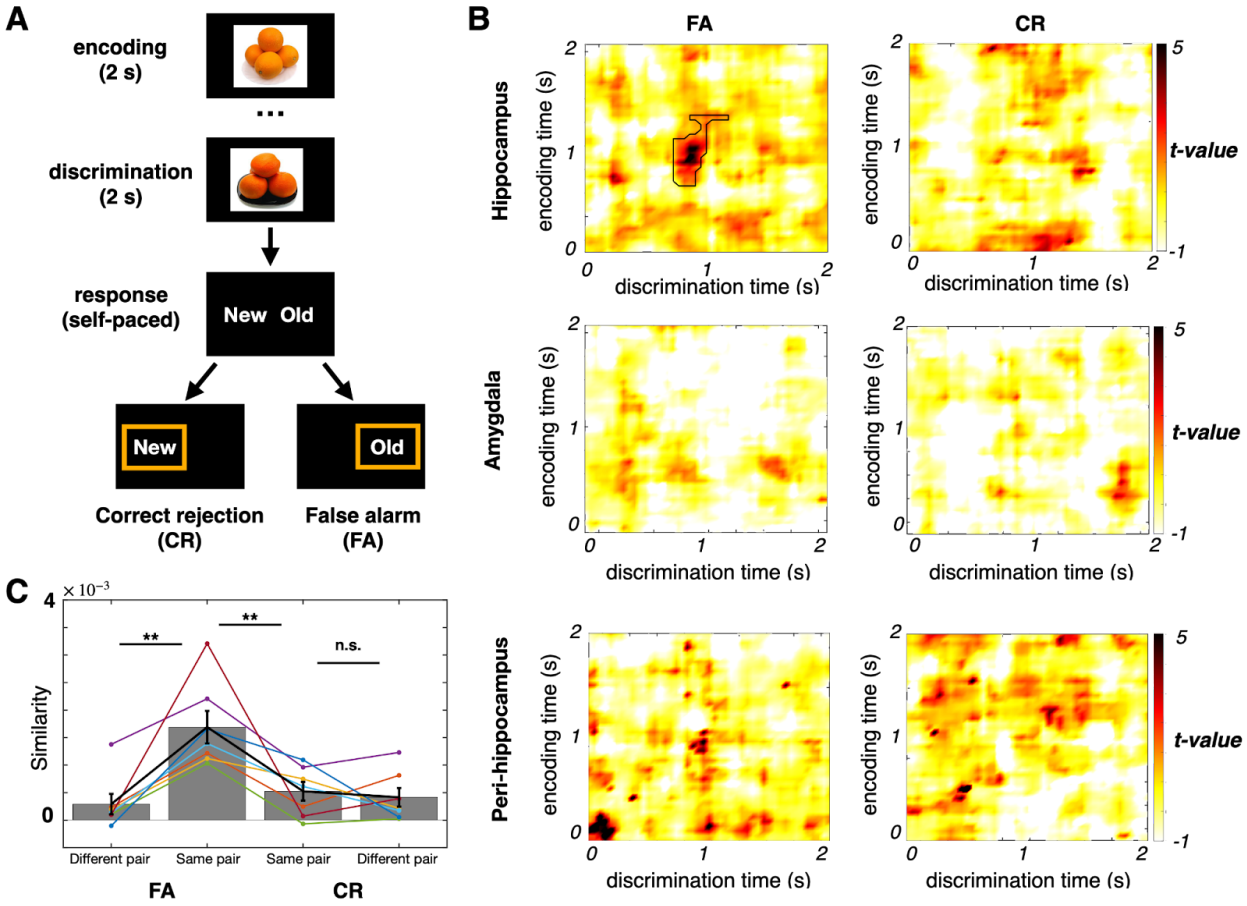


Fig. 3.3 Representational similarity to encoding trial predicts the correct discrimination of Lure stimuli.
a, Schematic view of Correct rejection (CR) trial and False alarm (FA) trial. **b,** T-value maps for the FA (Lure stimuli misclassified as Old, left) and CR condition (Lure stimuli correctly classified as Old, right), in the hippocampus, amygdala and parahippocampus. Significant representational similarity was present in the hippocampus during FA trials ($p = 0.042$, nonparametric cluster-based permutation test, encircled in black), suggesting the pattern completion as a neural mechanism underlying the misclassification of Lure stimuli as Old. **c,** Average hippocampal representational similarity (Spearman r) during the significant representational similarity window (encircled black in A). The colored lines represent

individual subjects. The Same-pair representational similarity was significantly higher during FA trials than Different-pair similarity ($p=0.002$, $t(6)=4.517$, one side paired t-test). There was no difference between the Same- and Different-pair similarity during CR trials ($p=0.32$, $t(6)=0.50$, one side paired t-test), while the Same-pair representational similarity was significantly lower during CR, relative to FA trials ($p=0.007$, $t(6)=3.356$, one side paired t-test), suggesting the pattern separation as a mechanism enabling the correct discrimination of highly similar Lure stimuli.

3.4.4 Subsecond network dynamics predicts the discrimination of highly similar stimuli

Pattern separation is a neural computation hypothesized to decrease memory interference by increasing the representational distance between the highly similar memories (O'Reilly and McClelland, 1994; Leutgeb et al., 2007; Favilla et al., 2016). Based on the notion of attractor-like organization of memory representations in the hippocampus (Hopfield et al., 1982; Morris and McNaughton, 1987), we predicted that the successful pattern separation would result in the formation of a distinct memory representation of a Lure stimulus. To test this hypothesis, we projected the neural spectral patterns from encoding and discrimination trials into a common low-dimensional state-space for each Target-Lure stimulus pair, using principal component analysis (Fig. 3.4a, b; Methods). Next, the k-nearest neighbor (k-NN) algorithm was applied to classify the discrimination trial time points as either: 1) Encoding state - within the neighborhood of the stimulus-specific encoding states, 2) Baseline state - the state within the neighborhood of the pre-trial baseline states, or 3) Outside state - the state outside either the Encoding or Baseline states (Fig. 3.4b). The average amount of time classified as Encoding state was significantly higher during FA

compared to CR discrimination trials, both in DG/CA3 and CA1 subfields ($p_{DG/CA3} = 0.040$, $p_{CA1} = 0.019$, paired t-test; Fig. 3.4c). The difference in the Encoding state visitation was reflected in the longer visitation to Outside state on CR trials in DG/CA3, but not in CA1 ($p_{DG/CA3} = 0.003$, $p_{CA1} > 0.05$, paired t-test; Fig. 3.4c). These results suggest that the state trajectory moving outside the neighborhood of previously encoded Target stimulus facilitates successful pattern separation. This result was consistent across the individual subjects (Fig. 3.4c). As expected, there were no conditional differences in the Baseline state visitation in either subfield ($p > 0.05$, paired t-test). These results further support the role of DG/CA3 in decreasing memory interference by pattern separation. Moreover, the flickering pattern of transition between the states is reminiscent of fast transitions between the competing hippocampal spatial maps in rodents (Jackson and Redish, 2007; Jezek et al., 2011).

Theoretical and empirical literature suggests that the hippocampal memory representations are organized as discrete attractors (Hopfield et al., 1982; Morris and McNaughton, 1987; Wills et al., 2005; Leutgeb et al., 2007; Jezek et al., 2011). If the Encoding and Outside sub-states represent distinct entities, they would be separated in the state-space, rather than a smooth transition. To test this hypothesis, we computed the velocity of state-space trajectory centered at the transitions between the Encoding and Outside sub-states (in either direction). State-space velocity is significantly higher at state transitions, relative to velocity null-distribution obtained by randomly shuffling the state identity ($p < 0.05$, see

Method 'State-space analyses' , Fig. 3.4d), consistent with the presence of distinct spatially-separated sub-states.

We hypothesized that the Lure stimulus presentation initially triggers the reinstatement of Target stimulus representation (Encoding state), followed by the later transition to Outside state. The prediction of such a dynamics is that the Encoding→Outside transition predicts the successful pattern separation. To test this hypothesis, we compared all the combinations of bidirectional state transitions between the CR and FA trials. Probability of Encoding→Outside transition was significantly higher during CR trials, compared to FA trials in DG/CA3, but not in CA1 hippocampal subfield ($p=0.006$, mixed-effect model, Fig. 3.4e). The transition probability difference was significant during the later part of the trial (~1-1.5 sec following the trial onset; $p<0.001$, nonparametric cluster-based permutation test, Fig. 3.4f), consistent with the initial retrieval of Encoding state, followed by the later transition into Outside state, associated with correct discrimination of similar stimuli (CR).

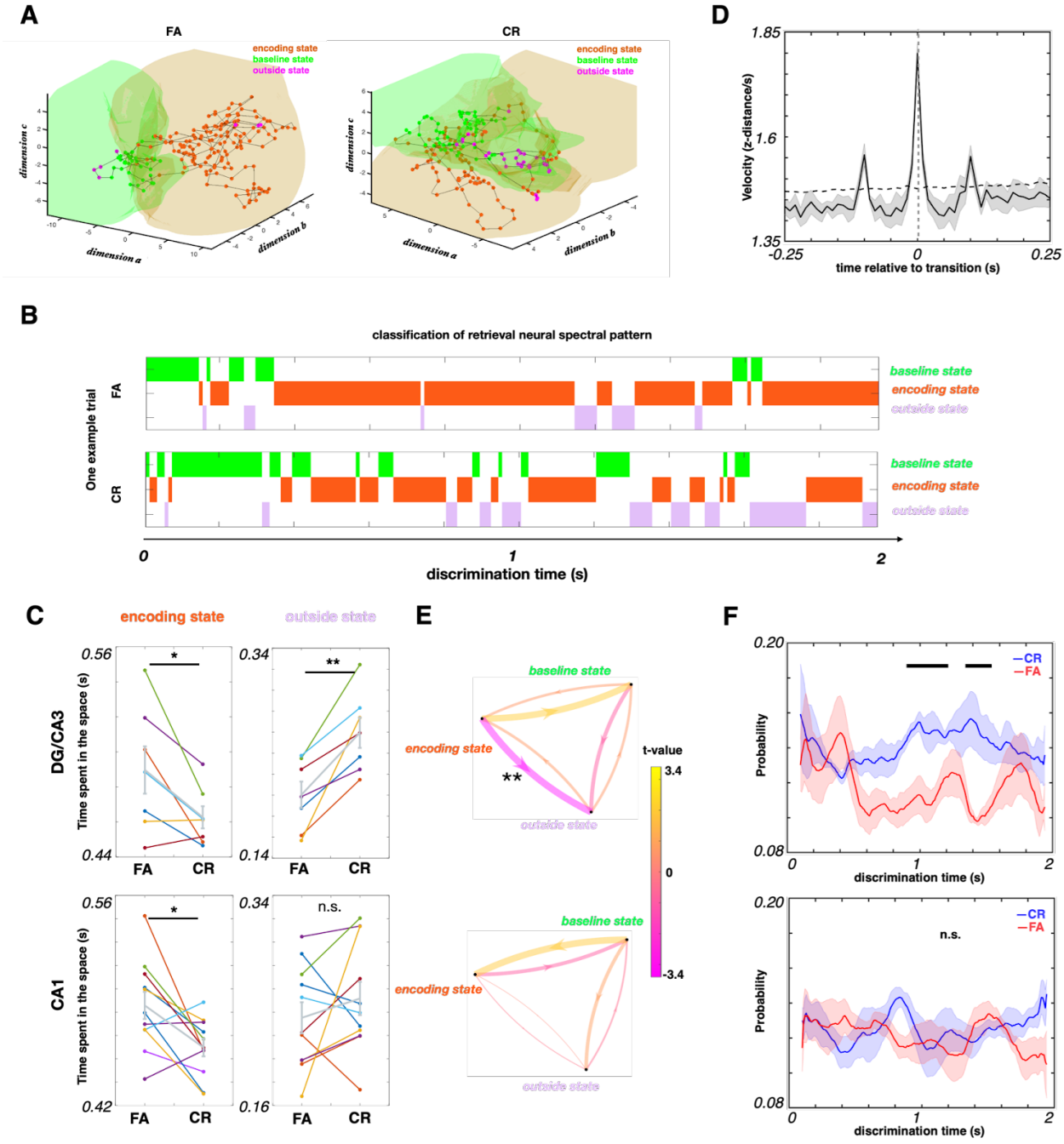


Fig. 3.4 Neural state-space dynamics in the hippocampus during Lure discrimination trials. Encoding state (0.5-1.5 s of the encoding trial) contains the stimulus-specific representation. Baseline state (pre-trial baseline on discrimination trial) is used to control for difference in temporal context. Neural activity during discrimination trials was classified as either Encoding or Baseline, using a k-nearest neighbour (k-NN) algorithm. Next, the time bins outside of Encoding or Baseline state radius (Methods) were classified as Outside state.
a, A state-space dynamics on an example FA (left) and CR (right) trials. Neural activity is projected into a common state-space (only the first three principal components are shown). The red and green manifolds represent the Encoding and Baseline state, respectively. The black line denotes a state-space trajectory. **b,** The linearized state-space trajectories from the example Lure

discrimination trials shown in (A). The state color coding is the same. Trajectory in the FA trials starts from Baseline state, transitions to Encoding state and mostly remains within the Encoding state for the rest of the trial. Similarly, the trajectory during CR trials starts from the Baseline state and transitions to the Encoding state, but later tends to settle in the Outside state. This dynamics is suggestive of pattern separation and predicts the correct discrimination. **c,** The total duration of state visitations for DG/CA3 (top) and CA1 (bottom) state-space trajectories, during CR and FA trials. Colored lines indicate individual electrodes and the gray lines indicate mean ± SEM. Both DG/CA3 and CA1 neural trajectory visits more Encoding state during FA trials ($p_{DG/CA3} = 0.040$, $p_{CA1} = 0.019$, paired t-test;). Only DG/CA3 activity spends more time in Outside state during the CR trials.($p_{DG/CA3} = 0.003$, $p_{CA1} > 0.05$, paired t-test;) **d,** State-space trajectory velocity abruptly increases at the state transitions, suggesting the large representational distance between the states, consistent with attractor-like organization of stimulus representations. Dashed line represents the velocity null-distribution, obtained by shuffling the temporal structure. **e,** Comparison of state transition probabilities between the CR and FA trials, for DG/CA3 (top) and CA1 (bottom). The arrow color and thickness depict the t and p-value (1 - p). Encoding to Outside transition probability is significantly higher on CR, relative to FA trials ($p=0.006$, mixed-effect model). **f,** Significantly higher transition probability from Encoding to Outside state is present during later part of CR discrimination trials (1.3 - 1.5 sec), relative to FA discrimination trials ($p<0.001$, nonparametric cluster-based permutation test). This suggests that the Lure stimulus presentation triggers the retrieval of encoded Target stimulus (Encoding state), followed by transition to Outside state on CR trials. Such an outcome leads to pattern separation and correct discrimination between highly similar stimuli. On the FA trials, the lack of transition to Outside state results in pattern completion and failure to discriminate between highly similar stimuli. However, such phenomena is not observed in CA1 ($p=0.11$, nonparametric cluster-based permutation test).

Expansion recoding predicts behavioral outcome

One of the hypothesized mechanisms of pattern separation is expansion recoding, defined as the increase in number of neurons representing a given memory between the hippocampal layers DG and CA3 (Marr 1971; Cayco-Gajic and Silver, 2019). Expansion recoding is predicted to increase the dimensionality of neural representations (Treves and Rolls 1994; Cayco-Gajic and Silver, 2019). To test this hypothesis, we quantified the dimensionality of neural signal during discrimination trials (Methods). The signal dimensionality increased early during discrimination trials, regardless of the behavioral outcome (Fig. 3.5A).

Dimensionality increase was continued throughout the CR trials, defined by the successful discrimination of highly similar stimuli. During FA trials, defined by the failed discrimination, dimensionality asymptoted ~1 sec following the trial onset. The slowing of dimensionality increase coincided with the temporal window of significant Target-Lure representational similarity on FA trials (Fig. 3.5a), indicated by an abrupt dimensionality increase drop (Fig. 3.5b). This is reflected by the significantly faster dimensionality increase on CR trials, 1600-1800 msec following the trial onset ($p = 0.008$, nonparametric cluster-based permutation test; Fig. 3.5b). To quantify the possible effect of representational similarity on dimensionality, we compared the dimensionality increase prior and after the representational similarity peak (Fig. 3.5a, dashed vertical line). Dimensionality increase during FA trials was significantly smaller after the representational similarity peak ($p = 0.002$, $t(3) = 5.06$, paired t-test, Fig. 3.5c left), suggesting that both measures might be reflecting the same underlying process, resulting in pattern completion and failure to discriminate the similar stimuli. This dynamics is consistent with the model whereby the dimensionality of Lure stimulus sensory information is increasing during propagation through the DG/CA3 network, due to divergent nature of network projections (Witter 1993). During FA trials, this activity triggers reinstatement of previously encoded Target stimulus representation, resulting in pattern completion, which stabilizes the activity within the Target stimulus attractor and attenuates the further dimensionality increase. On the CR trials, dimensionality keeps increasing,

resulting in the pattern separation and correct discrimination of Lure stimulus from the highly similar, previously encoded Target stimulus.

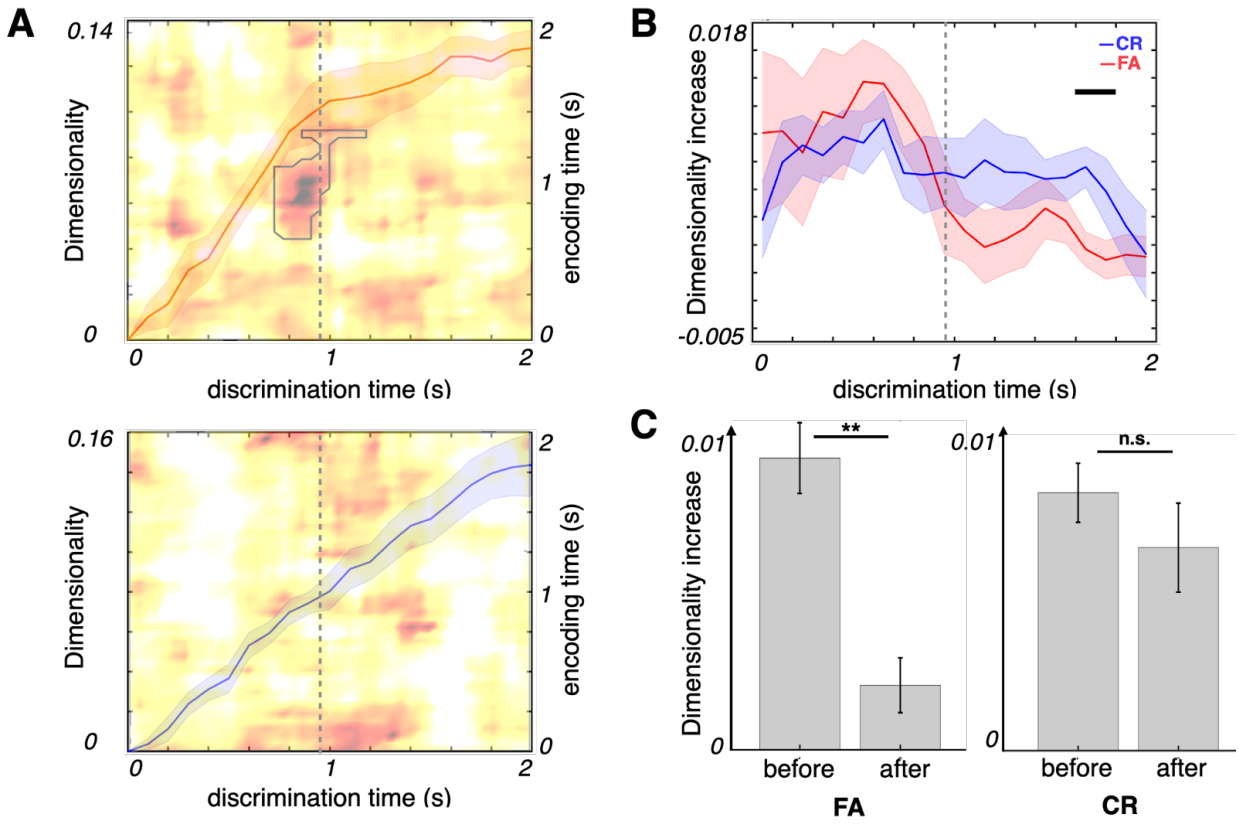


Fig. 3.5. Expansion recoding in DG/CA3 predicts the correct discrimination of highly similar stimuli.
a, The dimensionality of DG/CA3 neural activity during Lure discrimination trials (mean +/- SEM), superimposed on the representational similarity matrices from the same trials (Fig. 3B). FA trials (top): dimensionality asymptotes during and after the significant representational similarity window. CR trials (bottom): dimensionality increases throughout the discrimination trial. Dashed line - peak similarity timepoint. **b,** Dimensionality expansion (first derivative of dimensionality trace) is significantly higher on CR, compared to FA trials ($p = 0.008$, nonparametric cluster-based permutation test). **c,** Dimensionality expansion is significantly lower after the similarity peak on FA trials, but not different on CR trials (FA: $p = 0.002$, $t(3) = 5.06$; CR: $p = 0.344$, $t(3) = 1.027$, paired t-test).

3.5 Discussion

In the present study, we demonstrate the behaviorally relevant, sub-second pattern completion/separation dynamics in the human hippocampus. The behavioral relevance is reflected by the association between correct discrimination of highly similar stimuli (Target/Lure pairs) and degree of representational distance, neural state-space trajectory and representation dimensionality increase during discrimination trials. The observed dynamics is consistent with remapping as the reflection of underlying perceptual state (Sanders et al., 2020; Kubie et al., 2019; Fenton and Kelemen, 2016) and expansion recoding as a mechanism underlying pattern separation (Cayco-Gajic and Silver, 2019).

Pattern separation is defined as an input processing that results in increased representational distance between the non-simultaneously presented similar stimuli (Santoro, 2013; Madar et al., 2019). The opposite process, pattern completion, is defined as the reinstatement of previously encoded representation, following the presentation of an overlapping input (Hebb 1949; Treves and Rolls, 1994). Therefore, the demonstration of stimulus-specific representation in the human hippocampus is a necessary prerequisite for studying the pattern separation/completion dynamics. For this purpose, we demonstrated the presence of stimulus-specific hippocampal representation as the higher similarity between the multiple presentations of the same stimuli, relative to representations of randomly paired stimuli. The stimuli-specific

activity occurs ~1 sec following the stimuli onsets, suggesting the relatively long input processing in the hippocampal network, consistent with previous findings (Staresina et al., 2016). This finding aligns with previous reports based on fMRI, iEEG and single unit recordings (Lohnas et al., 2018; Gao et al., 2021; Gelbard-Sagiv et al., 2008, but see Pacheco-Esteban et al., 2019).

Pattern completion prevents mnemonic discrimination

Pattern completion is a process whereby the sensory input overlapping with previously encoded memory triggers memory reinstatement (Yassa and Stark, 2011). Extensively studied animal model of pattern completion consists of the exposure to environments with different degrees of similarity. This research suggests that the balance between the pattern completion/separation might reflect the animal's perception of the environment novelty (Miller and Sahay, 2019). By analogy, we hypothesized that the balance between the pattern completion/separation in the human hippocampus underlies mnemonic discrimination - ability to perceive the stimulus novelty, despite the high similarity to previously encoded stimulus. This hypothesis was tested by collecting the subjective perceptual report from human subjects, while recording the iEEG from temporal lobe structures, including the individual hippocampal subfields critical for pattern separation/completion. The misclassification of a new stimulus as previously experienced (false alarm - FA) is associated with significantly high representational similarity between the Target/Lure stimulus presentations (Fig 3.3a). Thus, the retrieval of previously encoded similar

stimulus representation, likely reflecting the pattern completion, is predictive of behavioral outcome - the conflation of two similar stimuli as identical. On the contrary, the Target/Lure representational similarity was not different from the randomly paired stimuli when the Lure stimulus was correctly perceived as novel (Fig. 3.3a). This suggests that the discrimination of similar stimuli is driven by the pattern separation, process of increasing the representational distance, consistent with hippocampal remapping. Behaviorally-relevant representational difference was present in the hippocampus, but not in other temporal lobe structures, consistent with hippocampal role in pattern completion/separation. These findings support the notion that hippocampal remapping reflects subjective perception of contextual change, rather than sensory input differences (Kelemen and Fenton,2010 ; Kubie et al., 2020; Sanders et al., 2020). Such an insight would be difficult to obtain in animal models, due to uncertain inference of the animal's hidden perceptual state.

Visitation of distinct states predicts the behavioral outcome

Rodent, non-human primate, and human fMRI studies suggest that the pattern separation is implemented in the DG/CA3 network (Wills et al., 2005; Leutgeb et al., 2007; Colgin et al., 2008; Jezek et al., 2011; van Dijk et al., 2018; Sakon and Suzuki, 2019; Julian and Doeller, 2020; Hainmueller et al., 2020; Wanjia et al., 2021). Several human fMRI experiments associated the perceptual change with hippocampal remapping, showing the abrupt change of multivariate pattern of hippocampal BOLD signal, coinciding with learning

milestones. While these studies convincingly point to pattern separation and hippocampal remapping as underlying the perception of contextual change, remapping was observed between the individual trials, due to temporal resolution constraint. To observe the temporal dynamics of remapping at individual trial level, we applied the state-space decoding approach, whereby the individual time bins during discrimination trials were classified based on the state-space distance to encoding trial. In this framework, the high or low proximity to the encoding trail is hypothesized to reflect the reinstatement of Target stimulus or the formation of new representation of Lure stimulus. In addition, to control for the change in temporal context as a source of state-space distance, the time bins with high proximity to pre-discrimination trial baseline were treated as a separate category. The total amount of time the neural trajectory spends in the Encoding or Outside state was predictive of behavioral outcome. This effect was present in DG/CA3, but not CA1 hippocampal subfield, consistent with lower correlation between the spatial maps of multiple environments in rodent DG/CA3 (Guzowski et al., 2004; Leutgeb et al., 2007). Encoding and Outside states don't represent a continuum in state-space, as reflected by the large state-space trajectory velocity at the state-transitions, suggesting a large separation between, relative to within the states. Such a state-space configuration suggests that the balance between the pattern completion/separation depends on the transition dynamic characteristic of distinct attractors. Temporal structure of state-transitions suggest that the network initially retrieves the Target stimulus representation, but the later

transitioning to Outside state - possibly reflecting the hippocampal remapping - predicts the behavioral outcome.The observed state-space trajectory is consistent with the retrieval of the most similar encoded spatial representation, when processing the ambiguous sensory input (e.g. Leutgeb et al., 2005). These results are in agreement with the observations of spatial maps switching on the theta cycle time scale in the rodent hippocampus (Jezek et al., 2011; Kelemen and Fenton, 2016). Specifically, the presence of additional velocity peaks ~200 ms from the state transition peak suggests that the state transitions occur at ~5 Hz. To our knowledge, this is the first demonstration of behaviorally relevant, subsecond time-scale representational switching in the human hippocampus. We also demonstrate the utility of the iEEG gamma band state-space approach in tracking the map dynamics at sub-second level (Jezek et al., 2011; Kelemen and Fenton, 2016; Kay et al., 2020; Farhoodi et al., 2020).

Remapping dynamics is consistent with expansion recoding

Expansion recoding is defined as increase in the number of neurons representing individual memories across the neural network layers, combined with non-linear mixing (Cayco-Gajic and Silver, 2019). This process is hypothesized as a mechanism underlying pattern separation (Marr, 1969; Albus, 1971; Treves and Rolls ,1998; Cayco-Gajic and Silver, 2019). Based on the prediction that expansion recoding results in increased dimensionality, we tested the signal dimensionality on discrimination trials. The initial dimensionality increase could reflect the input propagation through the DG/CA3 network,

characterized by fanning projections and huge increase in neuron number across the layers. The dimensionality asymptote on FA trials starts ~1 sec following the trial onset and coincides with the temporal window of high representational similarity between the Target/Lure trials. This dynamics is consistent with the Lure stimulus input triggering the pattern completion of the previously encoded Target stimulus representation, due to significant overlap of sensory information. On the CR trials, defined by the correct discrimination between the highly similar Target/Lure stimulus pair, dimensionality increased throughout the trial, possibly resulting in the formation of separate representation following pattern separation.

Chapter 4

Conclusion and Future Directions

4.1 Overview of the Findings

4.1.1 Hippocampal aSWR supports Enhanced Emotional Memory Formation

It is a long-held notion and common personal observation that emotional events are better remembered. Multiple lines of evidence have suggested neuromodulatory mechanisms on modifying plasticity during emotional memory encoding and offline sleep periods. Extensive studies suggest a key role of SWR in memory consolidation during sleep with emerging evidence for awake SWR (aSWR) occurrence associated with exposure to novel or reward-related contexts. However, it is largely unknown whether aSWR plays a role in emotional memory processing. Here we hypothesized that aSWR occurrence was associated with emotional experience, and coordinated amygdala-hippocampal activity, to support enhanced memory for the emotional event.

We found that the hippocampal aSWR occurrence during the immediate post-encoding period predicted both stimulus-induced arousal and subsequent memory performance. In addition, memory reinstatement was presented selectively during aSWRs, but not during post-encoding periods outside of

aSWR windows. Notably, memory discrimination was further predicted by the co-occurrence of aSWR-locked memory reinstatement in the amygdala and hippocampus, with the amygdala directionally influencing hippocampal memory reinstatement. These findings provide direct evidence linking aSWRs and emotional encoding through the engagement of amygdala-hippocampal networks.

4.1.2 Expansion Recoding supports Pattern Separation in the DG/CA3 Underlying Memory Discrimination

Discriminating a similar event from previously encountered one is a critical cognitive ability for adapting to novel situations. Pattern separation is a mechanism hypothesized to support the cognitive discrimination of similar representations in the brain. Theoretical works suggest pattern separation results in independent populations of neurons forming two distinctive attractors. We found that increased hippocampal representational similarity between the previously encoded and newly presented stimulus predicted failed discrimination, while decorrelated representational similarity predicts successful discrimination. Moreover, the hippocampal representational dynamic is consistent with discrete attractors, which are characterized by abrupt transitions between encoding and an outside state at a sub-second time scale.

A long-standing theory proposed expansion recoding as a putative mechanism underlying pattern separation. Anatomical studies suggest the DG

network as a neural substrate to implement expansion recoding by increasing neural activity space. We found the representational dimensionality in DG/CA3 was continuously increasing during successful discrimination, while a strong correlation with encoded memory representation impedes it from increasing during failed discrimination. Overall, these results support the discrete attractor hypothesis and suggest expansion recoding as a putative mechanism underlying pattern separation.

4.2 Future Directions

In the present study, we described the role of aSWR associated neural reinstatement during the post-encoding period in indexing the strength of memory encoding. Our findings show that formed cell ensembles representing the stimulus memory were reactivated and consolidated during the immediate post-encoding window (which occurred right after the active stimulus encoding period).

The two-stage model proposed that sequential activation during theta rhythm supports initial memory trace formation and subsequent sequential reactivation during SWR supports consolidation of memory representation (Buszaki, 1989). Recent finding report that place cell sequence underlying degraded theta rhythm when rodent was moved passively impaired subsequent SWR-associated reactivation during sleep, whereas theta-associated place cell sequence during active moving preserved the reactivation during sleep (Drieu et

al., 2018). This study constructed a causal relationship between theta and SWR sequences during wakefulness and sleep. However, an important question that remain is the relationship between theta-associated ensembles activation and aSWR-associated reactivation.

Although extensive studies suggest discontinuous human theta rhythms instead of the stable theta in the rodent hippocampus, theta power has been associated with the strength of memory encoding. Thus, it is plausible to make an analogy between rodent theta rhythm and human increased theta oscillatory power. Specifically, we can test the hypothesis that whether the neural representation occurred during elevated theta power window (i.e., theta-window) is predictive of the subsequent neural representation during aSWR window, compared to outside of the theta-window.

Accumulating evidence suggests that rapid eye movement (REM) sleep plays a critical role in strengthening and modulation of emotional memories (Yonelinas and Ritchey ,2015; Hutchison et al., 2015; Kensinger et al., 2009). Walker has proposed that REM sleep is implicated in two ways in emotional memory reprocessing: while the content of the memory is strengthened, the affective tone associated with the memory is weakened during REM sleep (Walker, 2009). However, physiological data supporting this proposal is lacking. Moreover, the distinctive roles of REM sleep and non-rapid eye movement sleep (NREM) sleep on emotional memory consolidation remains controversial. A recent finding showed preferential reactivation of fearful memory ensembles during non-rapid eye movement sleep (Girardeau et al., 2017), which suggests

NREM sleep is also involved in modifying emotional experience. In the future, we can investigate the occurrence of reinstatement during sleep. Specifically, whether the memory reinstatement will be observed during REM or NREM sleep. Moreover, investigating the temporal dynamics of memory reinstatement strength in the amygdala and hippocampus network during different instances of reinstatement events may shed light on testing the strengthening and weakening hypothesis of REM sleep (Walker, 2009).

REFERENCES

Aminoff, E. M., Kveraga, K., & Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends in cognitive sciences*, *17*(8), 379-390.

Amit, D. J., & Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. Cerebral Cortex, 7, 237–252.

Avants, B.B. et al. *Neuroimage* **54**, 2033–2044 (2011).

Axmacher, N., Henseler, M. M., Jensen, O., Weinreich, I., Elger, C. E., & Fell, J. (2010). Cross-frequency coupling supports multi-item working memory in the human hippocampus. *Proceedings of the National Academy of Sciences*, *107*(7), 3228-3233.

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *science*, *319*(5870), 1640-1642.

Bârzan, H., Ichim, A. M., & Mureşan, R. C. (2020, May). Machine Learning-Assisted Detection of Action Potentials in Extracellular Multi-Unit Recordings. In *2020 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)* (pp. 1-5). IEEE.

Battaglia, F. P., Sutherland, G. R., & McNaughton, B. L. (2004). Local sensory cues and place cell directionality: additional evidence of prospective coding in the hippocampus. *Journal of Neuroscience*, *24*(19), 4541-4550.

Ben-Yakov, A., & Henson, R. N. (2018). The hippocampal film editor: sensitivity and specificity to event boundaries in continuous experience. *Journal of Neuroscience*, *38*(47), 10057-10068.

Berron, D., Schütze, H., Maass, A., Cardenas-Blanco, A., Kuijf, H. J., Kumaran, D., & Düzel, E. (2016). Strong evidence for pattern separation in human dentate gyrus. *Journal of Neuroscience*, *36*(29), 7569-7579.

Bragin, A., Engel Jr, J., Wilson, C. L., Fried, I., & Buzsáki, G. (1999). High-frequency oscillations in human brain. *Hippocampus*, *9*(2), 137-142.

Buzsáki, G. (2015). Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus*, *25*(10), 1073-1188.

Buzsáki, G., & Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature neuroscience*, *16*(2), 130-138.

Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nature reviews neuroscience*, *13*(6), 407-420.
Cahill, L., & McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in neurosciences*, *21*(7), 294-299.
Canolty, R. T., & Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends in cognitive sciences*, *14*(11), 506-515.
Carr, M. F., Jadhav, S. P., & Frank, L. M. (2011). Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature neuroscience*, *14*(2), 147-153.
Carr, M. F., Karlsson, M. P., & Frank, L. M. (2012). Transient slow gamma synchrony underlies hippocampal memory replay. *Neuron*, *75*(4), 700-713.
Cayco-Gajic, N. A., & Silver, R. A. (2019). Re-evaluating circuit mechanisms underlying pattern separation. *Neuron*, *101*(4), 584-602.
Cohen, M.X. *MIT Press* (2014).
Colgin, L. L., & Moser, E. I. (2010). Gamma oscillations in the hippocampus. *Physiology*, *25*(5), 319-329.
Colgin, L. L., Moser, E. I., & Moser, M. B. (2008). Understanding memory through hippocampal remapping. *Trends in neurosciences*, *31*(9), 469-477.
Cover, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE transactions on electronic computers*, (3), 326-334.
Davachi, L., & Wagner, A. D. (2002). Hippocampal contributions to episodic encoding: insights from relational and item-based learning. *Journal of neurophysiology*, *88*(2), 982-990.
Dragoi, G. & Buzsaki, G. Temporal encoding of place sequences by hippocampal cell assemblies. Neuron 50, 145–157 (2006).
Drieu, C., Todorova, R., & Zugaro, M. (2018). Nested sequences of hippocampal assemblies during behavior support subsequent sleep replay. *Science*, *362*(6415), 675-679.
Druckmann, S., & Chklovskii, D. B. (2012). Neuronal circuits underlying persistent representations despite time varying activity. *Current Biology*, *22*(22), 2095-2103.
Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: spatial navigation and beyond. *Nature neuroscience*, *20*(11), 1504-1513.

Farhoodi, S., Plitt, M., Giocomo, L., & Eden, U. (2020). Estimating fluctuations in neural representations of uncertain environments. *Advances in Neural Information Processing Systems*, *33*, 10855-10868.

Folkerts, S., Rutishauser, U., & Howard, M. W. (2018). Human episodic memory retrieval is accompanied by a neural contiguity effect. *Journal of Neuroscience*, *38*(17), 4200-4211.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in cognitive sciences*, *9*(10), 474-480.

Fuhs, M. C., & Touretzky, D. S. (2007). Context learning in the rodent hippocampus. *Neural computation*, *19*(12), 3173-3215.

Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., and Moser, M.-B. (2004b). Spatial representation in the entorhinal cortex. *Science* 305, 1258–1264.

Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., & Fried, I. (2008). Internally generated reactivation of single neurons in human hippocampus during free recall. *Science*, *322*(5898), 96-101.

Genzel, L., Dragoi, G., Frank, L., Ganguly, K., De La Prida, L., Pfeiffer, B., & Robertson, E. (2020). A consensus statement: defining terms for reactivation analysis. *Philosophical Transactions of the Royal Society B*, *375*(1799), 20200001.

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, *31*(5), 681-697.

Girardeau, G., Inema, I., & Buzsáki, G. (2017). Reactivations of emotional memory in the hippocampus–amygdala system during sleep. *Nature neuroscience*, *20*(11), 1634-1642.

Guzman, S. J., Schlögl, A., Frotscher, M., & Jonas, P. (2016). Synaptic mechanisms of pattern completion in the hippocampal CA3 network. *Science*, *353*(6304), 1117-1123.

Guzowski, J. F., Knierim, J. J., & Moser, E. I. (2004). Ensemble dynamics of hippocampal regions CA3 and CA1. *Neuron*, *44*(4), 581-584.

Hainmueller, T., & Bartos, M. (2020). Dentate gyrus circuits for encoding, retrieval and discrimination of episodic memories. *Nature Reviews Neuroscience*, *21*(3), 153-168.

Hebb, D.O. (1949). The Organization of Behavior (New York: Wiley).

Helfrich, R.F. et al. *Nat. Commun.* **10**, 1–16 (2019).

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, *79*(8), 2554-2558.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of mathematical psychology*, *46*(3), 269-299.

Huang, N.E. et al. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **454**, 903–995 (1998).

Hutchison, I. C., & Rathore, S. (2015). The role of REM sleep theta activity in emotional memory. *Frontiers in psychology*, *6*, 1439.

Hyafil, Alexandre, et al. "Neural cross-frequency coupling: connecting architectures, mechanisms, and functions." Trends in neurosciences 38.11 (2015): 725-740.

Jackson, J., & Redish, A. D. (2007). Network dynamics of hippocampal cell-assemblies resemble multiple spatial maps within single tasks. *Hippocampus*, *17*(12), 1209-1229.

Jadhav, S. P., Rothschild, G., Roumis, D. K., & Frank, L. M. (2016). Coordinated excitation and inhibition of prefrontal ensembles during awake hippocampal sharp-wave ripple events. *Neuron*, *90*(1), 113-127.

Jezek, K., Henriksen, E. J., Treves, A., Moser, E. I., & Moser, M. B. (2011). Theta-paced flickering between place-cell maps in the hippocampus. *Nature*, *478*(7368), 246-249.

Johnson, A., & Redish, A. D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience*, *27*(45), 12176-12189.

Joo, H. R., & Frank, L. M. (2018). The hippocampal sharp wave–ripple in memory retrieval for immediate use and consolidation. *Nature Reviews Neuroscience*, *19*(12), 744-757.

Julian, J. B., & Doeller, C. F. (2021). Remapping and realignment in the human hippocampal formation predict context-dependent spatial behavior. *Nature Neuroscience*, *24*(6), 863-872.

Kelemen, E., & Fenton, A. A. (2010). Dynamic grouping of hippocampal neural activity during cognitive control of two spatial frames. *PLoS biology*, *8*(6), e1000403.

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion review*, *1*(2), 99-113.

Knierim, J. J., & Neunuebel, J. P. (2016). Tracking the flow of hippocampal computation: Pattern separation, pattern completion, and attractor dynamics. *Neurobiology of learning and memory*, *129*, 38-49.

Kondo H, Lavenex P, Amaral DG. Intrinsic connections of the macaque mon- key hippocampal formation: II. CA3 connections. The Journal of Comparative Neurology 2009;515:349–77.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., ... & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126-1141.

Kubie, J. L., Levy, E. R., & Fenton, A. A. (2020). Is hippocampal remapping the physiological basis for context?. *Hippocampus*, *30*(8), 851-864.

Leal, S. L., & Yassa, M. A. (2018). Integrating new findings and examining clinical applications of pattern separation. *Nature Neuroscience*, *21*(2), 163-173.

Leal, S. L., Tighe, S. K., & Yassa, M. A. (2014). Asymmetric effects of emotion on mnemonic interference. *Neurobiology of learning and memory*, *111*, 41-48.

Leutgeb, J. K., Leutgeb, S., Moser, M. B., & Moser, E. I. (2007). Pattern separation in the dentate gyrus and CA3 of the hippocampus. *science*, *315*(5814), 961-966.

Leutgeb, S., Leutgeb, J. K., Moser, E. I., & Moser, M. B. (2006). Fast rate coding in hippocampal CA3 cell ensembles. *Hippocampus*, *16*(9), 765-774.

Lever, C., Burton, S., Jeewajee, A., O'Keefe, J., & Burgess, N. (2009). Boundary vector cells in the subiculum of the hippocampal formation. *Journal of Neuroscience*, *29*(31), 9771-9777.

Lisman, J. E., & Jensen, O. (2013). The theta-gamma neural code. *Neuron*, *77*(6), 1002-1016.

Lisman, J., & Buzsáki, G. (2008). A neural coding scheme formed by the combined function of gamma and theta oscillations. *Schizophrenia bulletin*, *34*(5), 974-980.

Logothetis, N. K., Eschenko, O., Murayama, Y., Augath, M., Steudel, T., Evrard, H. C., ... & Oeltermann, A. (2012). Hippocampal–cortical interaction during periods of subcortical silence. *Nature*, *491*(7425), 547-553.

Lohnas, L.J. et al. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E7418–E7427 (2018).

Lopes-dos-Santos, V., van de Ven, G. M., Morley, A., Trouche, S., Campo-Urriza, N., & Dupret, D. (2018). Parsing hippocampal theta oscillations by nested spectral components during spatial exploration and memory-guided behavior. *Neuron*, *100*(4), 940-952.

Lum, J. A., Conti-Ramsden, G., Page, D., & Ullman, M. T. (2012). Working, declarative and procedural memory in specific language impairment. *cortex*, *48*(9), 1138-1154.

Lundqvist, Mikael, et al. "Gamma and beta bursts during working memory readout suggest roles in its volitional control." Nature communications 9.1 (2018): 394.

Lundqvist, Mikael, et al. "Gamma and beta bursts underlie working memory." Neuron 90.1 (2016): 152-164.

Lundqvist, Mikael, Pawel Herman, and Anders Lansner. "Theta and gamma power increases and alpha/beta power decreases with memory load in an attractor network model." Journal of cognitive neuroscience 23.10 (2011): 3008-3020.

Madar, A. D., Ewell, L. A., & Jones, M. V. (2019). Pattern separation of spiketrains in hippocampal neurons. *Scientific reports*, *9*(1), 1-20.

Marr, D. (1971). Simple memory: a theory for archicortex. Philos. Trans. R. Soc. Lond. B Biol. Sci. *262*, 23–81.

McGaugh, J. L. (2013). Making lasting memories: Remembering the significant. *Proceedings of the National Academy of Sciences*, *110*(Supplement 2), 10402-10407.

McGaugh, J. L. (2015). Consolidating memories. *Annual review of psychology*, *66*, 1-24.

McHugh, T. J., Jones, M. W., Quinn, J. J., Balthasar, N., Coppari, R., Elmquist, J. K., ... & Tonegawa, S. (2007). Dentate gyrus NMDA receptors mediate rapid pattern separation in the hippocampal network. *Science*, *317*(5834), 94-99.

McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D., & Eichenbaum, H. (2014). Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron*, *83*(1), 202-215.

McNaughton, B. L., & Morris, R. G. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in neurosciences*, *10*(10), 408-415.

Mehta, M. R., Lee, A. K., & Wilson, M. A. (2002). Role of experience and oscillations in transforming a rate code into a temporal code. *Nature*, *417*(6890), 741-746.

Miller, S. M., & Sahay, A. (2019). Functions of adult-born neurons in hippocampal memory interference and indexing. *Nature neuroscience*, *22*(10), 1565-1575.

Miyashita, Y. (2019). Perirhinal circuits for memory processing. *Nature Reviews Neuroscience*, *20*(10), 577-592.

Moca, V. V., Bârzan, H., Nagy-Dăbâcan, A. & Mureșan, R.C. *Nat. Commun.* **12**, 1–18 (2021).

Montgomery, S. M., & Buzsáki, G. (2007). Gamma oscillations dynamically couple hippocampal CA3 and CA1 regions during memory task performance. *Proceedings of the National Academy of Sciences*, *104*(36), 14495-14500.

Mormann, F., Fell, J., Axmacher, N., Weber, B., Lehnertz, K., Elger, C. E., & Fernández, G. (2005). Phase/amplitude reset and theta–gamma interaction in the human medial temporal lobe during a continuous word recognition memory task. *Hippocampus*, *15*(7), 890-900.

Neunuebel, J. P., & Knierim, J. J. (2014). CA3 retrieves coherent representations from degraded input: direct evidence for CA3 pattern completion and dentate gyrus pattern separation. *Neuron*, *81*(2), 416-427.

Neunuebel, J. P., Yoganarasimha, D., Rao, G., & Knierim, J. J. (2013). Conflicts between local and global spatial frameworks dissociate neural representations of the lateral and medial entorhinal cortex. *Journal of Neuroscience*, *33*(22), 9246-9258.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS computational biology*, *10*(4), e1003553.

Norman, Y. et al. *Science (80-.).* **365**, (2019).

O'reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus*, *4*(6), 661-682.

Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.M. *Comput. Intell. Neurosci.* **2011**, (2011).

Pacheco Estefan, D., Sánchez-Fibla, M., Duff, A., Principe, A., Rocamora, R., Zhang, H., ... & Verschure, P. F. (2019). Coordinated representational reinstatement in the human hippocampus and lateral temporal cortex during episodic memory retrieval. *Nature communications*, *10*(1), 1-13.

Peirce, J.W. *Front. Neuroinform.* **2**, 1–8 (2009).

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological review*, *116*(1), 129.

Popham, S. F., Huth, A. G., Bilenko, N. Y., Deniz, F., Gao, J. S., Nunez-Elizalde, A. O., & Gallant, J. L. (2021). Visual and linguistic semantic representations are aligned at the border of human visual cortex. *Nature Neuroscience*, *24*(11), 1628-1636.

Quian Quiroga, R. & Panzeri, S. *Nat. Rev. Neurosci.* **10**, 173–185 (2009).

Ramirez-Villegas, J. F., Logothetis, N. K., & Besserve, M. (2015). Diversity of sharp-wave–ripple LFP signatures reveals differentiated brain-wide dynamical events. *Proceedings of the National Academy of Sciences*, *112*(46), E6379-E6387.

Ray, S. & Maunsell, J.H.R. *PLoS Biol.* **9**, (2011).

Rennó-Costa, César, Daniel Garcia Teixeira, and Ivan Soltesz. "Regulation of gamma-frequency oscillation by feedforward inhibition: A computational modeling study." Hippocampus (2019).
Rolls, E. T., Treves, A., & Rolls, E. T. (1998). *Neural networks and brain function* (Vol. 572). Oxford: Oxford university press.
Rolls, E. T. (2010). A computational theory of episodic memory formation in the hippocampus. Behavioural brain research, 215(2), 180-196.
Rolls, E. T. (2016). Pattern separation, completion, and categorisation in the hippocampus and neocortex. *Neurobiology of learning and memory*, *129*, 4-28.
Sanders, H., Wilson, M. A., & Gershman, S. J. (2020). Hippocampal remapping as hidden state inference. *Elife*, *9*, e51140.
Santoro, A. (2013). Reassessing pattern separation in the dentate gyrus. *Frontiers in behavioral neuroscience*, *7*, 96.
Sauseng, P., Klimesch, W., Heise, K. F., Gruber, W. R., Holz, E., Karim, A. A., ... & Hummel, F. C. (2009). Brain oscillatory substrates of visual short-term memory capacity. *Current biology*, *19*(21), 1846-1852.
Shirvalkar, P. R., Rapp, P. R., & Shapiro, M. L. (2010). Bidirectional changes to hippocampal theta–gamma comodulation predict memory for recent spatial episodes. *Proceedings of the National Academy of Sciences*, *107*(15), 7054-7059.
Skelin, I., Zhang, H., Zheng, J., Ma, S., Mander, B. A., McManus, O. K., ... & Lin, J. J. (2021). Coupling between slow waves and sharp-wave ripples engages distributed neural activity during sleep in humans. *Proceedings of the National Academy of Sciences*, *118*(21).
Slutzky, M. W., Jordan, L. R., & Miller, L. E. (2008, August). Optimal spatial resolution of epidural and subdural electrode arrays for brain-machine interface applications. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 3771-3774). IEEE.
Sols, I., DuBrow, S., Davachi, L., & Fuentemilla, L. (2017). Event boundaries trigger rapid memory reinstatement of the prior events to promote their representation in long-term memory. *Current Biology*, *27*(22), 3499-3504.
Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological review*, *99*(2), 195.

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences*, *93*(24), 13515-13522.
Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annu. Rev. Neurosci.*, *27*, 279-306.
Staresina, B. P., Michelmann, S., Bonnefond, M., Jensen, O., Axmacher, N., & Fell, J. (2016). Hippocampal pattern completion is linked to gamma power increases and alpha power decreases during recollection. *Elife*, *5*, e17397.
Szőllősi, Á., & Racsmány, M. (2020). Enhanced mnemonic discrimination for emotional memories: the role of arousal in interference resolution. *Memory & cognition*, *48*(6), 1032-1045.
Talmi, D. (2013). Enhanced emotional memory: Cognitive and neural mechanisms. *Current Directions in Psychological Science*, *22*(6), 430-436.
Tolman EC. *Cognitive maps in rats and men.* American Psychological Association; 1948
Tort, A. B., Komorowski, R. W., Manns, J. R., Kopell, N. J., & Eichenbaum, H. (2009). Theta–gamma coupling increases during the learning of item–context associations. *Proceedings of the National Academy of Sciences*, *106*(49), 20942-20947.
Treves A, Rolls ET. A computational analysis of the role of the hippocampus in memory. Hippocampus 1994;4:374–91.
Trouche, S., Pompili, M. N., & Girardeau, G. (2020). The role of sleep in emotional processing: insights and unknowns from rodent research. *Current Opinion in Physiology*, *15*, 230-237.
Walker, M. P. (2009). The role of sleep in cognition and emotion. *Annals of the New York Academy of Sciences*, *1156*(1), 168-197.
Wang X-J, Buzsa´ki G. 1996. Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model. J. Neurosci. 16:6402–13
Wanjia, G., Favila, S. E., Kim, G., Molitor, R. J., & Kuhl, B. A. (2021). Abrupt hippocampal remapping signals resolution of memory interference. *Nature communications*, *12*(1), 1-11.
Whittington MA, Traub RD, Kopell N, Ermentrout B, Buhl EH. 2000. Inhibition-based rhythms: experimental and mathematical observations on network dynamics. Int. J. Psychophysiol.
Wills, T. J., Lever, C., Cacucci, F., Burgess, N., & O'Keefe, J. (2005). Attractor dynamics in the hippocampal representation of the local environment. *Science*, *308*(5723), 873-876.

Wiskott, L., Rasch, M. J., & Kempermann, G. (2006). A functional hypothesis for adult hippocampal neurogenesis: avoidance of catastrophic interference in the dentate gyrus. *Hippocampus*, *16*(3), 329-343.

Witter MP. Organization of the entorhinal–hippocampal system: a review of current anatomical data. Hippocampus 1993;3:33–44.

Wixted, J. T., Squire, L. R., Jang, Y., Papesh, M. H., Goldinger, S. D., Kuhn, J. R., ... & Steinmetz, P. N. (2014). Sparse and distributed coding of episodic memory in neurons of the human hippocampus. *Proceedings of the National Academy of Sciences*, *111*(26), 9621-9626.

Wixted, J.T. et al. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9621–9626 (2014).

Wu, C. T., Haggerty, D., Kemere, C., & Ji, D. (2017). Hippocampal awake replay in fear memory retrieval. *Nature neuroscience*, *20*(4), 571-580.

Wu, Z. & Huang, N.E. *Adv. Adapt. Data Anal.* **1**, 1–41 (2009).

Yaffe, R.B. et al. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18727–18732 (2014).

Yassa, M. A., & Stark, C. E. (2011). Pattern separation in the hippocampus. *Trends in neurosciences*, *34*(10), 515-525.

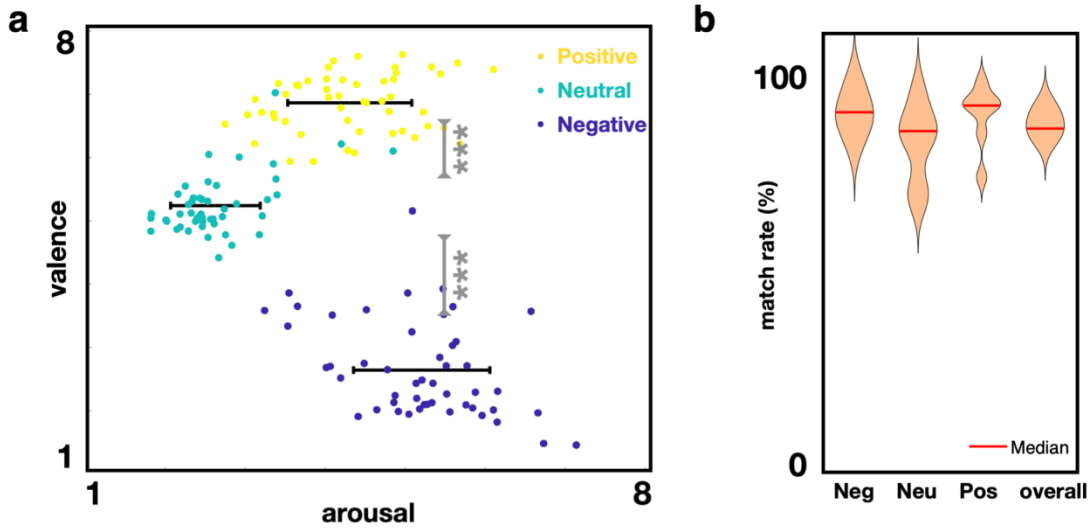
Yonelinas, A. P., & Ritchey, M. (2015). The slow forgetting of emotional episodic memories: An emotional binding account. *Trends in cognitive sciences*, *19*(5), 259-267.

Zhang, H., Fell, J. & Axmacher, N. *Nat. Commun.* **9**, (2018).

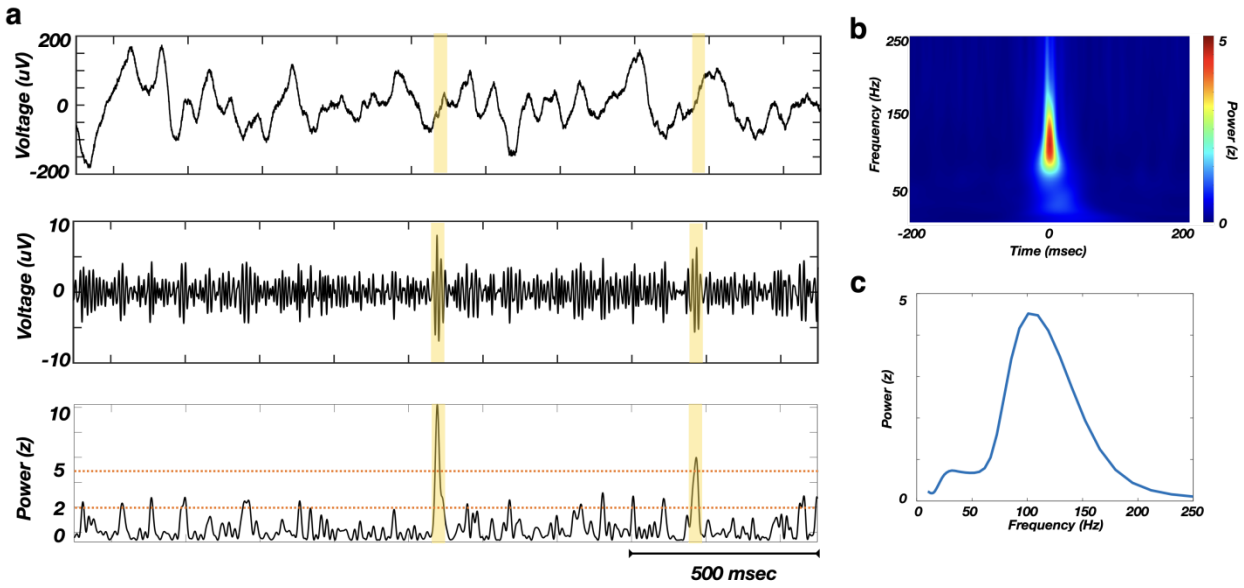
Zheng, J., Stevenson, R. F., Mander, B. A., Mnatsakanyan, L., Hsu, F. P., Vadera, S., ... & Lin, J. J. (2019). Multiplexing of theta and alpha rhythms in the amygdala-hippocampal circuit supports pattern separation of emotional information. *Neuron*, *102*(4), 887-898.

APPENDIX A

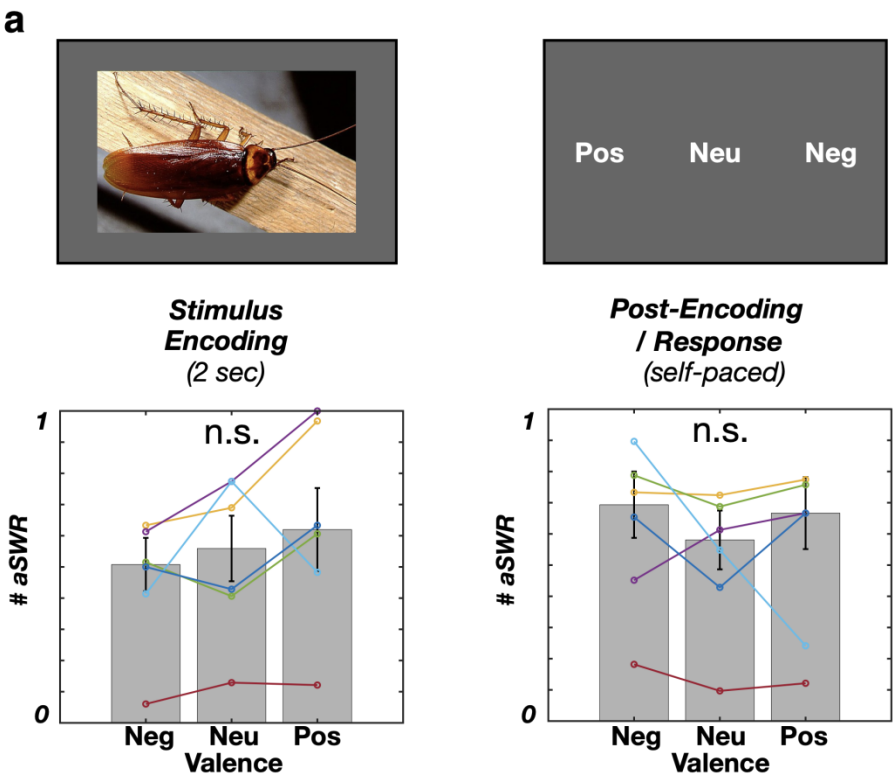
Supplemental Materials for Chapter 2



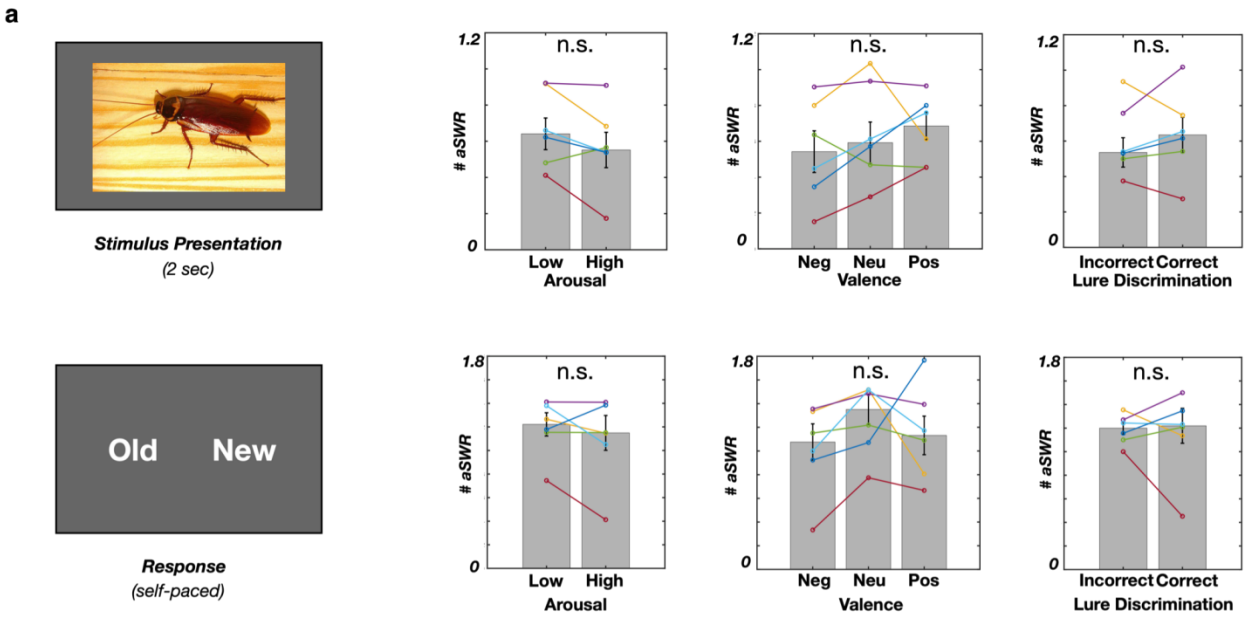
Supplemental Fig. 1. **a**, Positive and negative valenced stimuli are associated with higher stimulus-induced arousal, relative to neutral valence stimuli (***p<0.001, Wilcoxon rank-sum test). **b**, Stimuli valence ratings of study subjects are highly similar to the healthy population (match rate = 85.3 ± 1.3%).



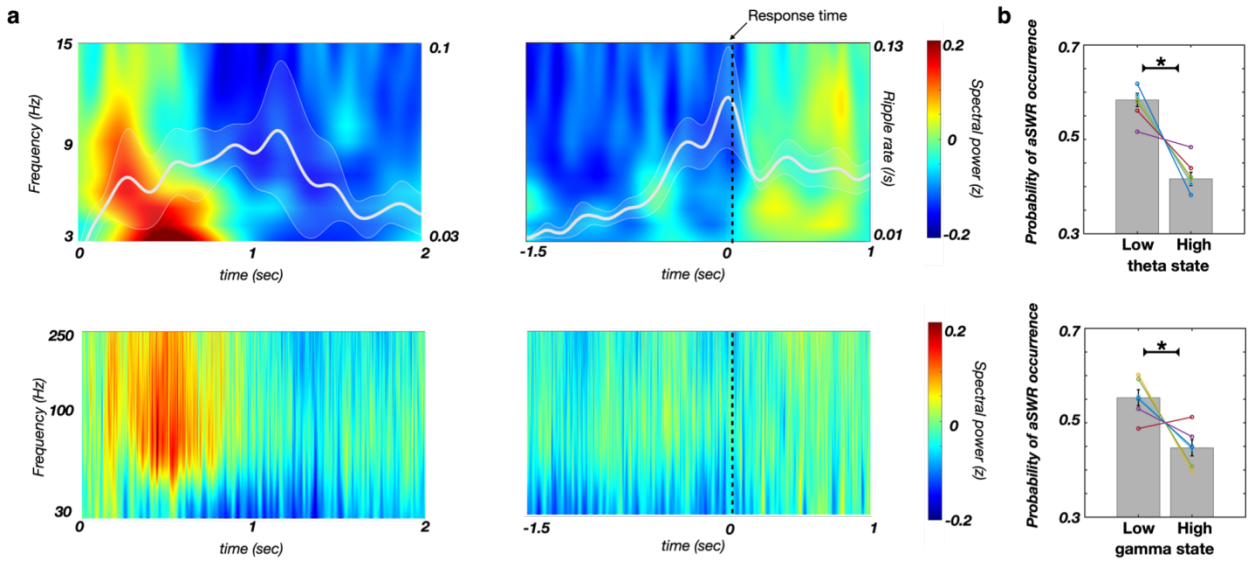
Supplemental Fig. 2. Awake SWR detection. a, Examples of several detected aSWRs (yellow highlights), showing the raw trace (top), filtered trace (80 - 150 Hz range, middle) and z-scored envelope of filtered trace (bottom). Detection is based on double-threshold (orange dashed lines) crossing of z-scored power (80-150 Hz) for the period of 20-100 msec. **b**, Z-scored power spectral density of average detected aSWR. **c**, Z-scored power during aSWR windows shows a bump in the 80-150 Hz range. This suggests that the aSWRs **are** not detected during signal artifact periods, which would reflect as a broadband power increase. In addition, detected aSWRs are not detected during non-specific increase in broadband gamma power or pathological high-frequency oscillations (> 200 Hz).



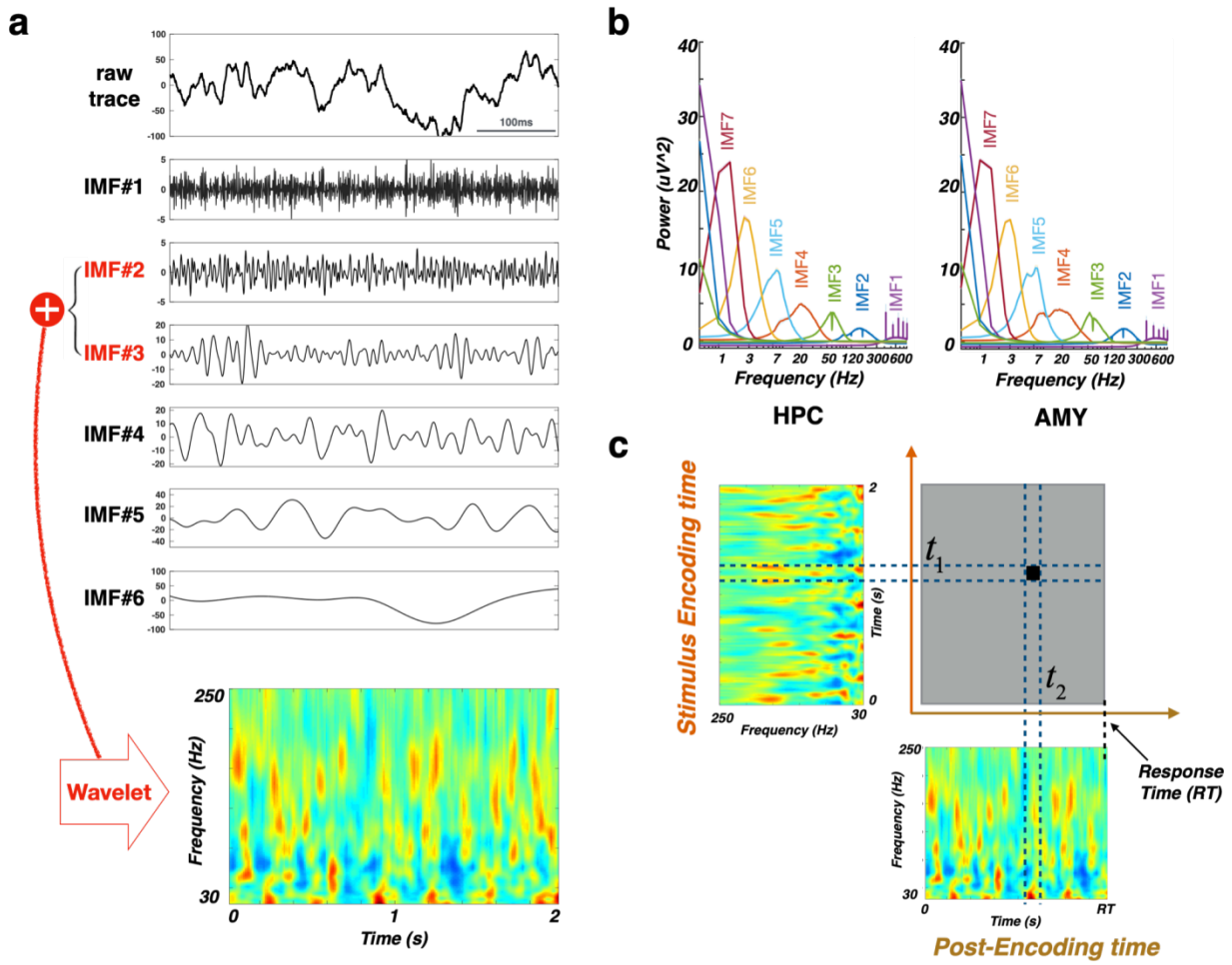
Supplemental Fig. 3. Stimulus valence is not significantly associated with aSWR occurrence during encoding stage. a, Stimulus encoding phase: $F(2, 15) = 0.67$, $p = 0.53$; Post-encoding: $F(2, 15) = 0.25$, $p = 0.77$, One-way ANOVA). The data from individual subjects are color-coded.



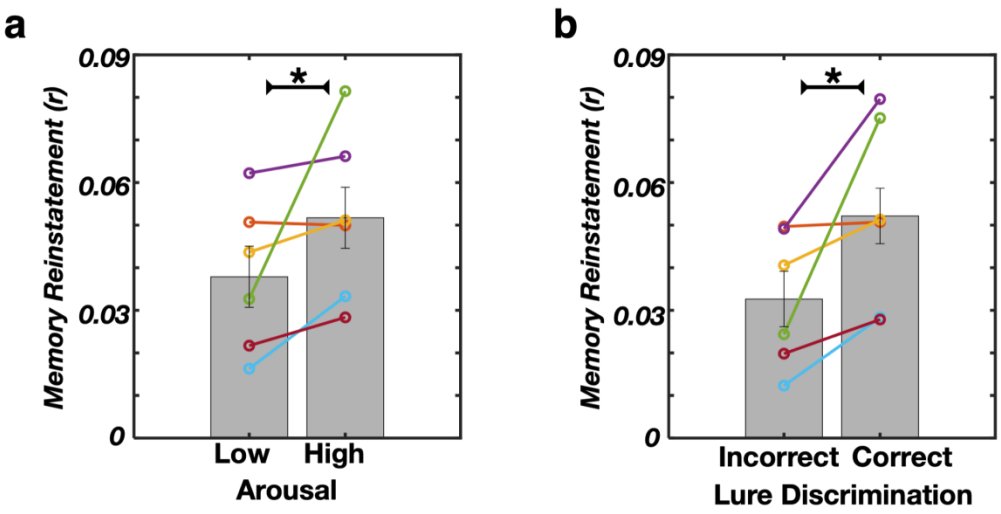
Supplemental Fig. 4. The aSWR occurence during retrieval task stage is not associated with stimulus-induced arousal, valence or correct discrimination. Arousal: Stimulus presentation (top row), $p = 0.11$, $z = 1.57$; Response (bottom row), $p = 0.17$, $z = 1.36$, Wilcoxon signed-rank test. **Valence**: Stimulus presentation, $p = 0.69$, $F(2, 15) = 0.69$; Response, $p = 0.51$, $F(2, 15) = 0.71$, One-way ANOVA). **Correct discrimination**: Stimulus presentation: $p = 0.6$, $z = -0.52$; Response: $p = 0.92$, $z = 0.11$, Wilcoxon signed-rank test).



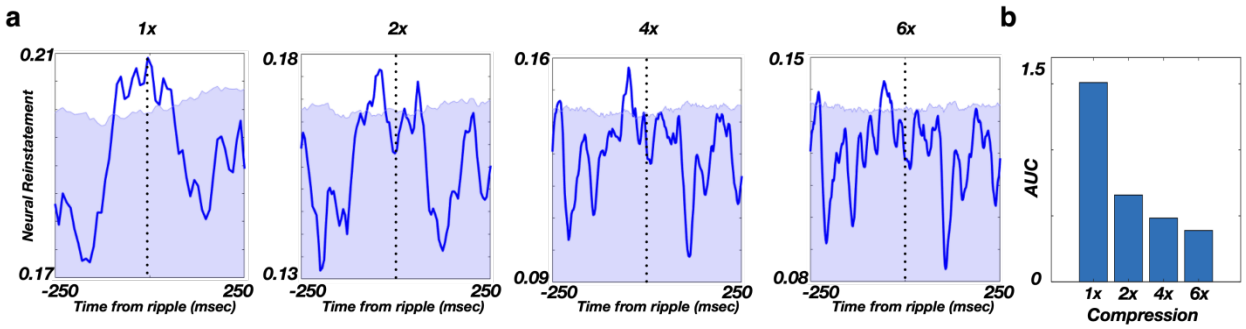
Supplemental Fig. 5. aSWRs occur predominately outside of high theta or broadband gamma periods. a, Low frequency (top, color) and high frequency spectrogram (bottom, color), and aSWR rate (white line) during the stimulus encoding (left) and post-encoding (right, response-locked) periods. **b,** The probability of aSWR occurrence is lower during the high theta state (top, $p = 0.017$, $z = 2.1$, one-tailed Wilcoxon signed-rank test), or during high gamma state (bottom, $p = 0.028$, $z = 1.9$, one-tailed Wilcoxon signed-rank test). Theta/gamma state classification was based on the power median split (for details, see 'Dual state analysis').



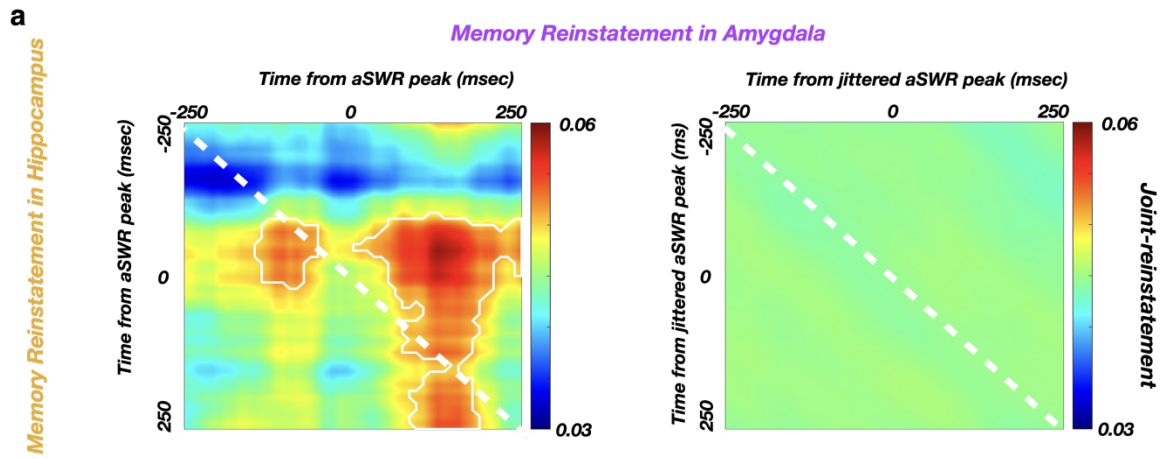
Supplemental Fig. 6. Overview of Ensemble Empirical Mode Decomposition (EEMD) and representational similarity analysis (RSA) methods. a, An example hippocampal raw iEEG trace (top) was decomposed into multiple intrinsic mode functions (IMFs; lower 6 panels). IMFs within the HFA range (IMF_2 and IMF_3) were used for HFA reconstruction. The HFA time-frequency matrix (bottom) was estimated using wavelet transformation (for details, see Time-frequency representation of the HFA). **b,** Power spectral density (mean ± SEM) of the IMFs decomposed from the hippocampal (left) and amygdala (right) electrodes. IMF spectral features were consistent across subjects and structures, with mean center frequencies in delta (IMF_7), theta (IMF_6, IMF_5), alpha/beta (IMF_4), gamma (IMF_3), high-gamma bands (IMF_2), and the noise term (IMF_1). The HFA time series were estimated by summing the IMFs with center frequencies > 30 Hz (IMF_2 and IMF_3). **c**, The similarity matrix (top right) was constructed by computing the power spectrum vector (PSV) Spearman's correlations for each combination of stimulus encoding (top left) and post-encoding (bottom right) time bins.



Supplemental Fig. 7. Post-encoding aSWR-locked reinstatement (amygdala and hippocampus combined) is increased for high stimulus-induced arousal and correctly discriminated stimuli. a, Arousal: *p = 0.046, z = -1.991, Wilcoxon signed-rank test. **b**, Correct discrimination: *p = 0.028, z = -2.201, Wilcoxon signed-rank test). Data from individual subjects is color-coded.



Supplemental Fig. 8. Post-encoding aSWR-locked memory reinstatement in the hippocampus is strongest without time compression. a, aSWR-locked hippocampal reinstatement during the post-encoding response period, across different temporal compression factors. Memory reinstatement strength area under curve (AUC) is defined as enclosed by the reinstatement trace (blue line) and 95% percentile of empirical null-distribution (blue shading upper limit). AUC reflects the memory reinstatement strength at different compression factors. **b**, Memory reinstatement strength is highest with no compression.



Supplemental Fig. 9. Joint cross-structure memory reinstatement occurs selectively during aSWR time windows. a, Average joint cross-structure reinstatement (hippocampus and amygdala) relative to aSWR peak times (left) and relative to jittered aSWR peak times (right). The white line encircles the periods of significant joint cross-structure memory reinstatement (Fig. 2.3d). The color scale represents the Spearman correlation between the encoding stimulus presentation and post-encoding aSWR windows. The absence of significant joint cross-structure memory reinstatement following the jittering of aSWR peak times (right) reveals the specificity of cross-structure reinstatement to aSWR windows.

Subject	Gender	Age
S1	M	21
S2	F	58
S3	M	24
S4	F	55
S5	M	23
S6	M	29
S7	F	21

Supplemental Table 1. Demographic information for the study subjects.

	Hippocampus	Amygdala
IMF#1	420 Hz	420 Hz
IMF#2	159 Hz	184 Hz
IMF#3	57.5 Hz	53 Hz
IMF#4	20 Hz	18 Hz
IMF#5	8 Hz	7.5 Hz
IMF#6	2.5 Hz	3 Hz
IMF#7	1.5 Hz	1 Hz
IMF#8	0.5 Hz	0.5 Hz
IMF#9	< 0.5 Hz	< 0.5 Hz
IMF#10	< 0.5 Hz	< 0.5 Hz

Supplemental Table 2. Center frequencies of the IMFs in the hippocampus and amygdala.

APPENDIX B

Supplemental Materials for Chapter 3

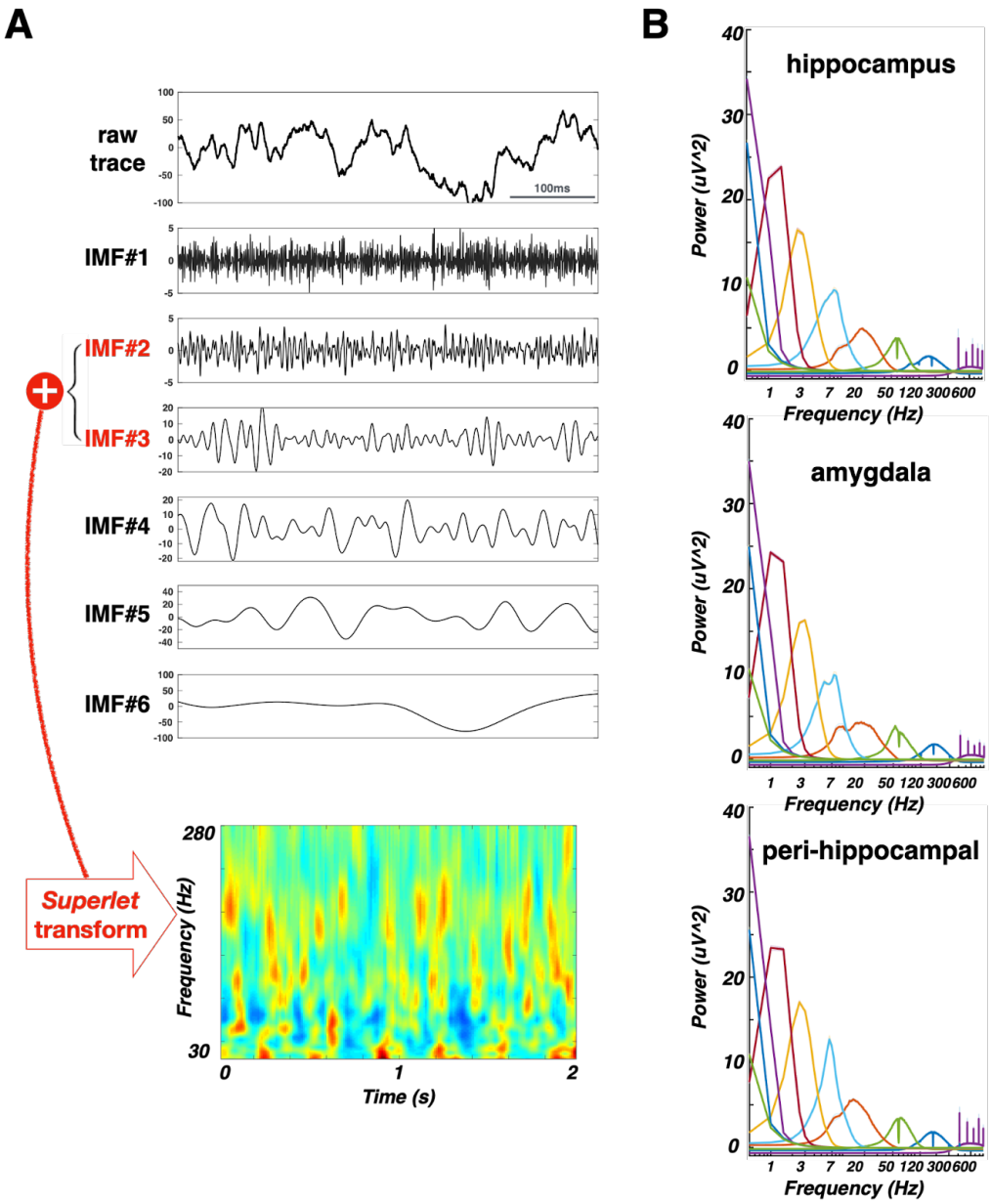


Fig. S1. Overview of the ensemble empirical mode decomposition (EEMD). a, The raw iEEG (top panel) was decomposed into multiple intrinsic mode functions (IMFs 1-6; lower 6 panels). IMFs with peak frequencies within the gamma range (30 - 300 Hz; IMF_2 and IMF_3) were used for gamma activity reconstruction. **b,** Power spectral density (mean ± SEM) of the IMFs decomposed from the hippocampal (left), amygdala (middle) and peri-hippocampal (right)

electrodes. Spectral features were consistent across the electrodes and anatomical structures. The mean center frequencies correspond to delta (IMF_7), theta (IMF_6, IMF_5), alpha/beta (IMF_4), gamma (IMF_3), high-gamma bands (IMF_2), and the noise term (IMF_1). The gamma time series are the sum of the IMFs with center frequencies > 30 Hz (IMF_2 and IMF_3).

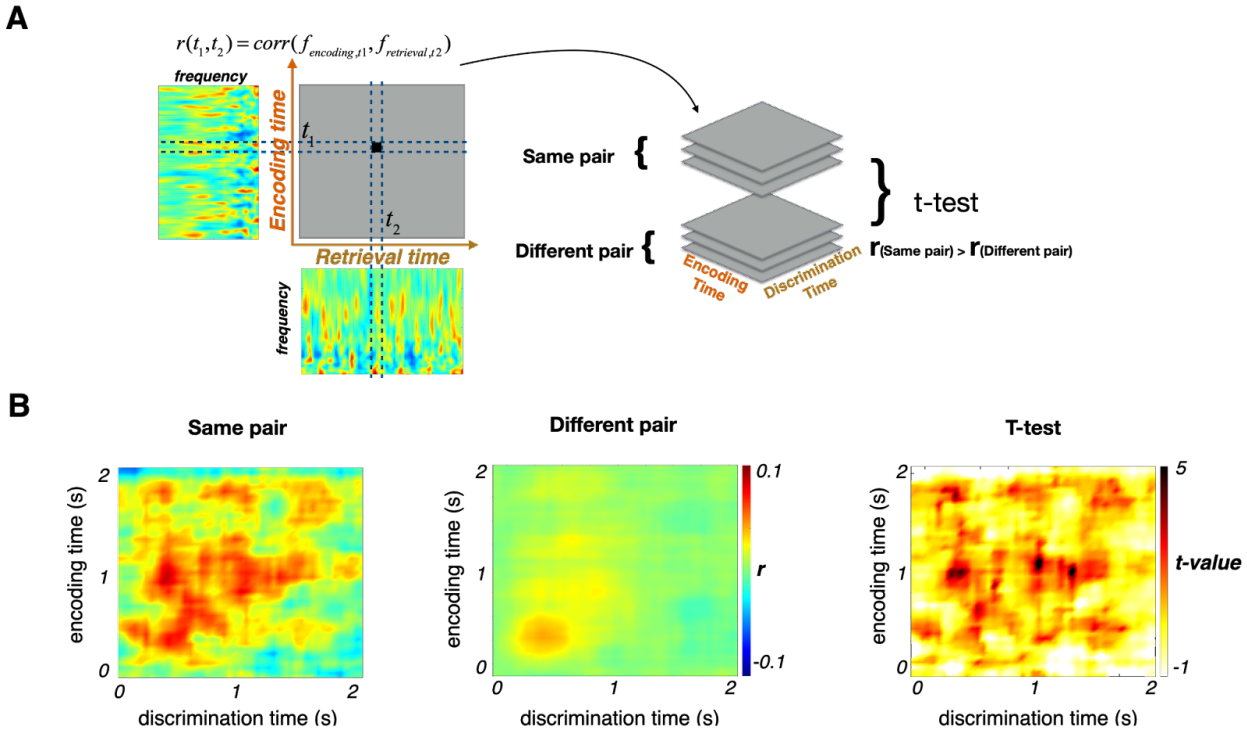


Fig. S2. Representational similarity analysis (RSA). a, Representational similarity matrix was obtained by computing the Spearman correlations between the power spectrum vectors for each combination of encoding and discrimination trial time bins (left). Significance was assessed by comparing the similarity of the same stimulus trial pairs (Same) with encoding - discrimination trial pairs of randomly paired stimuli (Different), using two-tail t-test and non-parametric cluster-statistics with 1000 permutations ($p < 0.05$; Fig. 2). The analysis in Fig. 3 is identical, except that the Target-Lure trial pairs were compared with randomly paired stimuli. **b,** Representational similarity matrices (left and middle) and t-value matrix (right) obtained using the procedure described in **A**.

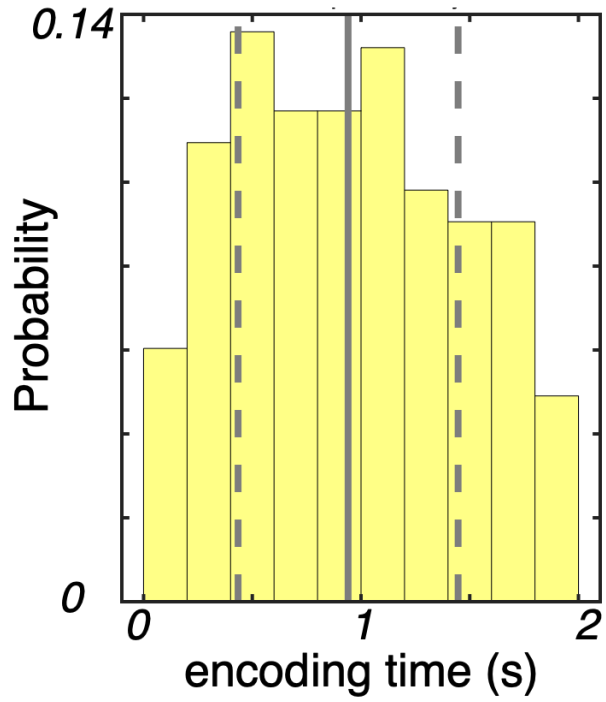


Fig. S3. Distribution of peak representational similarity timing across individual Repeat trials. The median +/- SD (the thick and dashed vertical lines) of this distribution was used to define the Encoding space window..

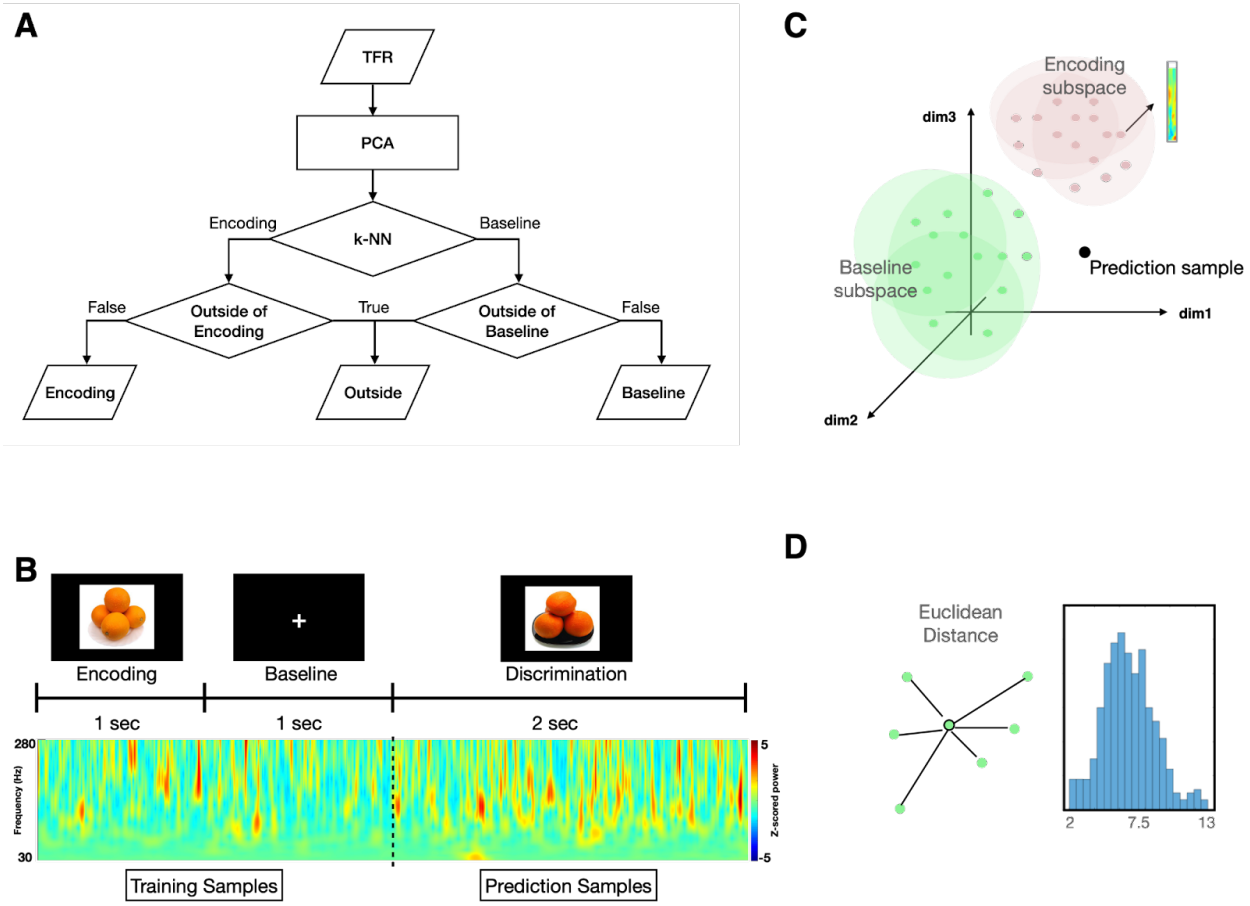


Fig. S4. State-space analysis overview. a, Flow diagram of the designed algorithm. **b,** Following the gamma trace reconstruction (Fig. S1), gamma trace TFR was computed using the wavelet transform. Next, the TFRs from the encoding trial (median similarity peak ± SD, duration 1000 ms; Fig. S3), pre-discrimination trial baseline (1000 ms) and discrimination trial (2000 ms) were concatenated to produce the encoding-discrimination pair-specific TFR. **c,** Dimensionality reduction was performed on pair-specific TFR using principal component analysis (1) and the first n PCs with cumulative explained variance (EV) > 80% were projected on a TFR matrix, resulting in a projection matrix.. Each projection matrix time bin was defined by an n-dimensional vector, denoting the respective PC weights (PC vector). Next, the pairwise Euclidean distances were computed between each timepoint PC vector and all other PC vectors from the same epoch (encoding or baseline) and the Encoding or Baseline sub-state boundaries were defined as the union of all the encoding or baseline boundaries. Next, the K-nearest neighbor (K-NN) algorithm was used to classify the PC vectors from discrimination trial based on the Euclidean distance to Encoding or Baseline sub-states (2). The PC vectors outside the 95th percentile of either the Encoding or Baseline substrate were classified as Outside sub-state (3). **d,** Pairwise distance distribution were computed and the 95th percentile were identified as radius for each points.

Subject	Gender	Age
S1	M	21
S2	F	58
S3	M	24
S4	F	55
S5	M	56
S6	M	23
S7	M	29
S8	F	21

Table 1. The demographics of study subjects.