UNIVERSITY OF CALIFORNIA,
IRVINE

System theory-based investigation of the brain: From Behavior Across Species to Solving a Fundamental Neural Code

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILISOPHY

In Electrical Engineering

By

Sandra Gattas

Dissertation Committee:
Professor Michael Yassa, PhD
Professor Lee Swindlehurst, PhD
Assistant Professor Yanning Shen, PhD

2021

DEDICATION

I dedicate this thesis to my mother Samira Eskandr and to my ultimate provider Jesus Christ.

TABLE OF CONTENTS

page

LIST OF FIGURES

LIST OF TABLES

LIST OF ILLUSTRATIONS

ACKNOWLEDGEMENTS

I am indebted to individuals that have planted seeds throughout my life critical in shaping who I am today and the work produced here.

Frist, I am grateful for my PhD advisor, Professor Michael Yassa. Professor Yassa has opened numerous and the most different of doors for me, enabling me to become the scientist I am. I could not have imagined thriving and succeeding to this quality and in such a short time with any other mentor. From day 1, when I knew little to today, he empowered me. His empowerment of others is invariant to seeing outcomes. Through this point of empowerment, he taught me many things—neuroscience content, logistical matters about academic processes, how to care for students as a supervisor, grant writing, executing fun, caring and successful collaborations, industry, developing and pursuing transformative visions and more. As he watched me master each lesson, he taught the next only by demonstrating it himself and in the most fun of ways. With his mentorship I am here today – completing a dissertation I am pleased with and have clarity and confidence in the line of work for my career. I have grown as a scientist and as an individual. Professor Yassa is a lifelong mentor and friend.

I am grateful for my co-adivsor Professor Gary Lynch who supported my ideas and me as a person. He opened his lab for me to test an idea that was seemingly impossible. Today, we celebrate its success. Throughout this journey, we enjoyed our time, especially our brainstorming sessions, where he taught me from his history of neuroscience knowledge and inquisitively engaged in engineering-based discussions with me.

I am also indebted to the UCI Electrical Engineering and Computer Science (EECS) department. Despite no prior background in electrical engineering, coming in and requesting to be a graduate student in EECS, Professor Swindlehurst didn't discourage my seemingly impossible request and rather treated me as a colleague and became my mentor; our discussions were inquisitive, fun, and educated me on how to rigorously link electrical engineering and neuroscience. I am also inspired by Professor Yanning Shen and I'm grateful for the invaluable insight and guidance she's provided me. Along the way, I matured into an Electrical Engineer with their mentorship and the training offered by EECS. Pursuing my PhD in EECS has been one of the most exciting journeys in my life. I want to thank the department and my course Professors Hamid Jafarkhani, Paniz Ibrahimi, Syed Jafar, Zak Kassas, Homayoun

Yousefi'zadeh, Keyue Smedley, and Mohamad Abdel Latif for the rigorous coursework put forth that gave me opportunities to wrestle with concepts day and night. I will miss these days deeply.

I thank my neuroscience co-mentors. First, Dr. Jack Lin - I grew from his mentorship; his joy in teaching and mentoring me and advocation for me and my career empowers me. Professor Bruce McNaughton for believing in me and my work; he recognized the value of my invention of the obstacle Course and scaled it across laboratories. His mentoring me as early as I was in medical school, shaped how I (and the field) think about the brain early on. Professor Majid Mohajerani for his faithful mentorship and opportunity for collaboration; I was proud to be part of his scholarly work which ended up shaping my own work. Professor Boyer Winters, for his mentorship and immense support in recruiting his skilled team to rigorously validate my invention of the Obstacle Course. Finally, Professor Josef Parvizi, my first neuroscience mentor, who saw my potential when I knew nothing about the brain nor signal processing. He is my inspiration to become a physician-scientist and opened this door for me by teaching me and being my biggest fan.

Another gratitude to funding support from the National Institute of Health Medical Scientist Training Program (NIH/MSTP) award (T32GM008620), pre-doctoral fellowship from the National Institute of Neurological Disorders and Stroke (NINDS/NIH) training grant (T32NS45540) and National Research Science foundation (NRSA) individual fellowship award from the National Institute of Aging (NIA/NIH, F30AG069406) which enabled my training and research.

I am deeply grateful for my undergraduate institution, San Jose State University, and its irreplaceable professors who challenged me into becoming a diligent student and thinker and nourished the budding scientist and dancer in me from age 16-21 years old; Professor Daryl Eggers, my first research advisor, and Professors Nemes Peter, Shelley Cargill, Katie Wilkinson, Maureen Scharberg, and William Murray.

My teachers in dance throughout my life have nourished my creativity, love for life, and identification and appreciation of beaty and patterns. I'm indebted to my undergraduate dance professors Gary Masters and Fred Mathews, and high school ballet teachers Deanna Ross and Laura Akard.

I'm grateful for my middle school, Pacific Grove Middle School, that welcomed me in the United States when I moved from Egypt at age 11. I have a lasting memory of my math teacher, Ms. Demarco; despite cultural and language barriers, she saw me, my talent in mathematics and nourished my love for math at 12 years old.

In addition to formal teachers, mentors and professors, I've grown and learned from my lab colleagues. I am grateful for Amy Daitch and Jen Chun Hsiang who mentored me in my first neuroscience lab at Stanford. These engineers inspired me to become an engineer. I am also grateful to my neuroscience collaborates Aliza Le, Heather Collette, Ethan Huff and Samantha Creighton and Gabriel Elias for their hard work as collaborators on the presented work and to Conor Cox, Rebecca Stevenson and Logan Harriger for helpful and insightful discussions.

I thank my friends that have been like family to me for as long as I can remember. Tunt Manal who loved and mentored me since childhood, Barbora Mistry my good friend since middle school and Jenny Duong my college roommate and good friend. I thank my Canadian family, Leah Parker and Sue Maclean, Nolette MacDonald and Renee Scotland, and Lisa Vike and Ravali Kosaraju, who cheered me on during my PhD and through their love, gave me a sense of home whenever I was in Canada—whether for collaborations or leisure.

I'm grateful for my home country Egypt where I learned how to become a diligent hard worker, understood the value of education, and developed a love for math and science at young age. I also thank Egypt for the memories it gave me; Egypt provided me a childhood well lived.

I want to thank my two sisters, Marina Parker and Sylvia Gattas, for seeing me, loving me and cheering me on for as long as I can remember. They were there during the fun and difficult times. The loved me on my good days and were graceful to me on my bad days. I can count on their love – my sisters are the definition of lifelong friends.

There is an individual that I grew with as a person and a thinker – this is Javad Karimi. He supported me during difficult times in medical and graduate school. When I was new to engineering, Javad listened to my questions, carefully thought through them, and discussed answers with me. The discussions he and I had were deeply intellectually stimulating, productive and purely enjoyable. I cherish them. They tapped greatly into my intellectual potential – planting a large intellectual seed. I'm indebted to his mentorship and collaboration.

I'm indebted to my mother who faithfully sacrificially and unconditionally loved and supported me. She worked hard to put me in the best school in the middle east, Ramses College for Girls, where I grew intellectually during my formative years. She was the first to plant the seed of math and science in me at a young age. I remember the day she taught me to write number 2 and how proud she was of me. She worked two jobs to provide for me and my sisters. She is my angel.

Any natural talents I have, where I grew up and immigrated to, and the people that saw me, loved me, supported me and invested in me along the way are gifts. I thank my ultimate provider, Jesus Christ, for these gifts.

VITA

Sandra Gattas

EDUCATION

MD **current**

University of California, Irvine

School of Medicine, Medical Scientist Training Program

PhD in Electrical Engineering **2021**

University of California, Irvine

MS in Electrical Engineering **2021**

University of California, Irvine

BS in Biology **2013**

San Josef State University

BA in Dance **2013**

San Josef State University

RESEARCH EXPERIENCE

Graduate Research Student **2017-2021**

University of California, Irvine

Key mentors: Michael Yassa, Gary Lynch, Lee Swindlehurst, Jack Lin, Bruce McNaughton

Key Collaborations: Majid Mohajerani (University of Lethbridge), Boyer Winters (University of Guelph), Casey Halpern (Stanford University)

Research Assistant and Laboratory Manager **2013-2015**

Stanford University, Josef Parvizi

Undergraduate research assistant **2011-2013**

San Jose State University, Daryl Eggers

REFEREED JOURNAL PUBLICATIONS

Huang, Y., Kakusa, B., Feng, A., **Gattas, S.**, Shivacharan, R., Lee, E., Parker, J. J., Kuijper F. M., Barbosa, D. A, Keller, C. J, Bohon, C., Abanoub, M., and Halpern, C. H. (2021). **The insulo-opercular cortex encodes food-specific content under controlled and naturalistic conditions.** *Nature Communications*.

Choi, E. H., **Gattas, S.,** Brown, N., Limbo, J. N., Chan, A. Y., and Oh, M. Y. (2021). **Epidural Electrical Stimulation for Spinal Cord Injury.** *Neural Regeneration Research.*

Kakusa, B., Huang, Y., Barbosa, D. A, B., Feng, A., **Gattas, S.**, Shivacharan, R., Lee, E., Parker, Kuijper F. M., Saluja, S., Parker, J. J., Miller, K. J., Keller, C. J., Bohon, C., and Halpern, C. H. (2021). **Anticipatory human subthalamic area beta-band power responses to dissociable tastes correlate with weight gain.**
Neurobiology of Disease, 154(105348).

Karimi Abadchi, J., Nazari-Ahangarkolaee, M., **Gattas, S.**, Bermudez-Contreras, E., Luczak, A., McNaughton, B. L., & Mohajerani, M. H. (2020**). Spatiotemporal patterns of neocortical activity around hippocampal sharp-wave ripples.** *ELife*, *9*.

Kakusa, B., Saluja, S., Dadey, D., Barbosa, D., **Gattas, S.,** Miller, K. J., Cowan, R. P., Kouyoumdjian, Z., Pouratian, N., & Halpern, C. H. (2020). **Electrophysiology and Structural Connectivity of the Posterior Hypothalamic Region: Much to Learn From a Rare Indication of Deep Brain Stimulation.** *Frontiers in human neuroscience*, *14*, 164.

Chan, A. Y., Lien, B. V., Choi, E. H., Chan, A. K., Hanna, G., Lopez, A., Brown, N., **Gattas, S.,** Kirillova, L., Horton, D., Fote, G., Hanst, B., Perry, R., Lee, Y., Golshani, K., Hsu, F. P., and Oh, M. Y. (2020). **Back Pain Outcomes After Minimally Invasive Anterior Lumbar Interbody Fusion: a Systematic Review.** *Journal of Neurosurgery.*

Lien, B. V., Brown, N. J., **Gattas, S.,** Choi, E. H., Sahyouni, R., Campos, J. K., Zhang, A., & Oh, M. Y. (2020). **The market landscape of online second opinion services for spine surgery.** *Surgical neurology international*, *11*, 365.

Brown, N., Lien, B. V., Shahrestani, S., Choi, E. H., Tran, K., **Gattas, S.,** Ransom, S. C., Tafreshi, A. R., Ransom, R. C., Sahyouni, R., Chan, A., and Oh, M. Y. (2021). **Getting Down to the Bare Bones: Does laminoplasty or laminectomy With Fusion Provide Better Outcomes for Patients With Multilevel Cervical Spondylotic Myelopathy?** Neurospine, 18(1).

Daitch, A. L., Foster, B. L., Schrouff, J., Rangarajan, V., Kaşikçi, I., **Gattas, S.**, & Parvizi, J. (2016). **Mapping human temporal and parietal neuronal population activity and functional coupling during mathematical cognition.** *Proceedings of the National Academy of Sciences of the United States of America*, *113*(46).

MANUSCRIPTS ACCEPTED, IN REVISION OR SUBMITTED

Gattas, S., Fote, G., Brown, N., Lien, B., Choi, E. H., Chan, A. Y., Rosen, C. D., Oh, M. **Second opinion in spine surgery: a scoping review.** (accepted, *Surgical Neurology International)*

Gattas. S, Collett, H. A., Huff, E. A., Creighton, S. D., Weber, S.E., Buckhalter, S., Manning, S. A., B. D. Winters and B. L. McNaughton. **The obstacle course: a novel enrichment paradigm for rodents**.

Gattas, S., Elias, G. A., Janecek, J., Yassa, M. A., & Fortin, N. (2020). **A role for proximal CA1 20-40 Hz dynamics for performance on a nonspatial sequence memory task.**

Gattas. S, Larson, M. S., Mnatsakanyan, L., Sen-Gupta, I., Vadera, S., Swindlehurst, L., Rapp, P., Lin, J. J., and Yassa, M. A. **Theta-mediated human hippocampal-neocortical interactions underlie pattern separation.**

Gattas, S., Le, A. A., Karimi Abadchi, J., Pruess B., Shen, Y., Swindlehurst, L., Yassa, M. A., Lynch, G. S. **Solving the fundamental neural code of the hippocampal CA1 apical dendritic system.**

MANUSCRIPTS READY FOR SUBMISSION

Gattas, S.*, Barbosa, D. N.*, Salgado, J. S., Kuijper, M. F., Huang, Y., Halpern, C. H. **Elucidating the feeding circuit in the human hippocampus.** *Authors contributed equally to this work.

PUBLISHED ABSTRACTS

Gattas, S, Huang, Y., Kakusa, B., Feng, A., Barbosa, D. A, Shivacharan, R., Parker, J. J., Kuijper F. M., C. J, Bohon, C., Abanoub, M., and Halpern, C. H. (2020). **Insular gamma power is sufficient to classify between anticipation of palatable versus neutral items.** *Neurosurgery.*

Sandra Gattas, Myra Larson, Jack J. Lin, and Michael A. Yassa. **Pattern separation beyond the hippocampus: neocortico-hippocampal mechanisms of pattern separation in humans.** Chicago, Illinois, Society for Neuroscience, 2019.

PATENTS

Neural control signals for therapeutic behavioral modulation, U.S. Patent Application No. 63/170,404 (provisional)

Inventors: Halpern C., Buch V., Shivacharan R., Gattas S.

Assigned to Stanford University

HONORS AND AWARDS

F30 National Research Service Award, 2020-2021, 2022-2024

T32 NIH training award, 2019-2020

Jared M. Roberts Memorial Award, 2020

Best Data Blitz presentation award, 32nd Annual Scientific Conference of the Center for the Neurobiology of Learning and Memory, Irvine, California, 2019

Honor standing (top 10% in course) in Neuroscience and Biomedical ethics, UCI School of Medicine, 2016

1st Place, PLEXUS Journal of Arts and Humanities, Performing Arts Competition, 2016

Publication of contemporary dance performance, 'When It's Time' in the PLEXUS Journal of Arts and Humanities, 2016 edition, UCI School of Medicine.

ABSTRACT OF THE DISSERTATION

System theory-based investigation of the brain: From Behavior Across Species to Solving a Fundamental Neural Code

By

Sandra Gattas

Doctor of Philosophy in Electrical Engineering

University of California, Irvine, 2021

Professors Michael Yassa and Lee Swindlehurst, Co-Chairs

Neuroscience is the study of the brain. It is an interdisciplinary field. At one level, behavioral tasks are carefully designed to understand behavior and cognition. At other levels, different devices and technologies are used to acquire neural data (the brain communicates via electrical signals) at different scales and link such data to behavior. The interdisciplinary nature is not just evident in the data acquisition but also in the processing and analyses of the data. Signal processing and other system theory-based techniques, statistical methods and computational modeling can be used to identify useful information in brain-acquired signals (electrical, blood flow, volume, etc.) and link them to behavior and cognition in healthy and diseased states.

As a field, neuroscience has made numerous meaningful discoveries primarily using the traditional hypothesis driven, top-down, deductive approach. In this approach, the goal is to

identify local phenomena that can account for macroscopic (behavioral) effects. This approach provides answers to the specific questions or hypotheses tested. My early work follows this approach.

In my first deductive-approach study, I developed a behavioral task that enables controlled delivery of rodent enrichment. I invented an Obstacle Course, which is a track with built-in obstacles that rodents must overcome to complete a lap and receive reward. Comparing this track with traditional enrichment (toys in larger home-cages) we found that animals running this track outperformed animals receiving traditional enrichment. This track then improves upon the rodent model for enrichment, which is used to test the hypothesis that enriched human life experience plays a protective role against myriad debilitating neurological diseases including dementia.

In the second deductive-approach study, I analyzed data from surgically implanted depth electrode recordings of electrical brain activity in rats engaged in a behavioral task where they were tested on their memory for a sequence of odors (memory for temporal order). Spectral analysis on this data revealed that the hippocampal area CA1 (a region important for memory processing) recruits power in the 20-40 Hz range during odor-processing, and that the steady-state dynamics of this neural synchrony occurs after a decision is made about temporal order. 20-40 Hz power during this post-decision state increases with knowledge of the odor-sequence. Lastly, machine learning analysis revealed that power in this post-decision state is predictive of sequence memory response accuracy. 20-40 Hz power is generally suggested to gate information within a brain network implicated in a given cognitive process (suppress distractors and maintain representations in facilitating sensori-motor integration); here we extend its role to sequence memory processing in the hippocampus.

The third deductive study was similar in nature to the aforementioned one, but in humans. Neural electrical signals were recorded from the brains of patients suffering from epilepsy, using surgically implanted electrodes for clinical monitoring, while they engaged in a

memory task. Spectral analysis revealed increased 4-5 Hz power in the hippocampus and neocortex, two key memory systems in the brain, during successful discrimination between similar memories. In this work, it was essential to make inferences about the presence and directionality of interactions between the two memory systems recruiting 4-5 Hz power. This led to my fourth deductive project -- expansion of two methodologies; 1) measuring 4-5 Hz phase transfer entropy but as a function of time, and 2) using the Kalman filter to estimate parameters of a time-varying autoregressive model for subsequent Granger causality analyses – both methods enabled inferences about the direction of communication between the two memory systems. Using these methods, I identified a pattern of directional communication between the hippocampus and neocortex that is consistent with predictions from computational models.

These deductive works led to new discoveries. However, I encountered a boundary point in deductive research – I can only know the answers to the specific hypotheses tested. But my knowledge about the general operational range and dynamics of the underlying brain circuits was absent. Given the astounding complexity of the brain, and the numerous behavioral phenomena it supports, deductive approaches alone do not suffice, as they may leave us with incomplete and perhaps misleading information regarding the underlying brain circuits. So I changed my approach to an inductive one.

An inductive, bottom-up approach, which is conspicuously absent in the field, reveals physiological rules that link biological phenomena with behavior. I therefore employed Volterra series-based system identification as an inductive approach and comprehensively, agnostically, and rigorously identified the operations of a given brain area, the hippocampal CA1. This was implemented by stimulating and recording from rodent brain slices that were kept live through an artificial physiological environment. Nearly the entirety of the operational range of the CA1 is captured in a single formula, an input-output nonlinear transfer function. Using this function alone, the CA1 output to a random input in a new set of animals can be predicted with high accuracy. Moreover, examining this formula led to new discoveries and provided insight about

the CA1 system software, despite this being one of the most heavily studied nodes in the brain. Contrary to the fields popular belief, the formula revealed the presence of system nonlinearity in CA1. Additionally, intuition is gained by examining this formula regarding how the CA1 treats inputs (amplification, suppression, harmonic generation) based on the content of the input.

All in all, across species, and using different modalities and approaches I summarize work concerned with first identifying specific neural electrical activity for specific behavior. Then I transition to a different approach -- identifying the fundamental rules of a candidate brain area, capturing nearly the entirety of its processes and operations in a single input-output transfer function. Potential future work could extend this approach to other brain nodes and link them together appropriately based on known anatomical connectivity. An exciting possible outcome of such a research program would be to ultimately characterize the entire brain circuit, through iterative system identification, such that a physically realizable brain can be built.

Chapter 1

Introduction

Neuroscience is the study of the nervous system, which includes the study of the brain. The fundamental cell for communication in the brain is a neuron (there are other cell types essential for the health and function of the brain). Neurons in the brain communicate via electrical signals. Therefore, the brain is a large electrical circuit, however, with some topology. Certain anatomical areas in this large circuit (subcircuits or nodes) have clear functions – language, motor, visual processing centers, etc.

The structure (hardware) and function (software) of the brain circuit have been studied for decades. This has been done through an interdisciplinary approach. Advanced technologies and equipment are used for data acquisition; electrophysiology (electrode insertion in the brain to record electrical activity), magnetic resonance imaging (MRI) to measure brain structure or blood flow patterns (functional MRI, fMRI), magnetic resonance spectroscopy (to measure certain metabolites), and many more techniques. Similarly, the interdisciplinary nature is evident on the data analysis end. Signal processing may be applied to extract useful information from neural data and link it to cognition and behavior. Statistical concepts are applied for significance testing, machine learning and model generation. Mathematics and physics are critical for generating theories or computational models regarding brain operations at the level of a neuron to a larger scale brain sub-circuit.

Over the last decades, the neuroscience field has made meaningful progress regarding brain structure and function. Briefly discussed are some examples of progress in the neuroscience field. With sufficient certainty, different brain areas are ascribed to functions (sensory, motor, language, memory, emotion, reward processing, etc.) and the anatomical connectivity between these areas are also well identified. Properties of certain cell types

(channels, neurotransmitters, cell morphology, etc.) are also well mapped. In fact, there are over 27,000 pubmed articles to the search term ‘neuroscience’.

The traditional approach taken in neuroscience studies is a ‘top-down’ deductive approach. In this approach, the goal is to identify local phenomena that can account for macroscopic (behavioral) effects. This provides answers to the specific questions or hypotheses tested. For example, lesion studies can be used to test if the absence of one brain area leads to impairment in a particular behavior. Or, by measuring neural activity during a particular behavior, an association between such activity and its possible function can be made. In this approach, hypotheses can be supported and possibilities can be narrowed down. However, only that which is precisely examined and tested is revealed – absent all other untested questions or even unimagined possibilities.

Despite this tremendous progress, there remains a critical gap. This gap is that the transfer dynamics that give rise to the fundamental unit of communication in the brain is unknown – how electrical spike patterns of a presynaptic cell are processed and transformed into output spike patterns by the post synaptic cell (the synapse is the junction of communication between a pair of neurons). Until this transformative process is identified, we are limited in generating an adequate theory for how the brain works, and more so, ever building one.

A natural consequence of this lack of knowledge, a lack of system identification of the brain circuits and subcircuits, is a lack of knowledge regarding deviation from normal (disease). There is a tremendous amount of debilitating neurological and neuropsychiatric disorders that are not defined (depression, schizophrenia, autism, dementia, etc.) – only a subset of observable symptoms and/or pathological markers are associated with these diseases. Such collection of symptoms and pathological observations are all the tools employed clinically to guide diagnosis and treatment. However, if brain circuits are fully mapped (identified), then in theory, one can identify the deviations in such mapping during disease, thereby defining the

mappings and operations of the disease circuitry. In this sense, the generality of the disease, its definition, would be the metric for diagnosis and drug development.

Conspicuously absent from the strategies of research employed in neuroscience are complementary 'bottom-up' inductive approaches through which identification of fundamental principles can be attained. Secondary to attaining such revealed fundamental rules and principles would be to link biological phenomena with behavior. In other words, identifying the behavioral consequences that may arise because of the identified basic circuitry and its operations.

System identification provides a means to both employing inductive strategies in neuroscience and addressing the critical gap of identifying the transformative process of input to output electrical spike patterns. System identification is concerned with complete characterization on an unknown system (i.e., a black box; a physical hardware for which its operations are unknown). These unknown operations can be summarized in a single mathematical formula – an input-output transformation function. Identifying the system means knowing its transformation function, which reveals how the system operates on an arbitrary input to generate an output. In the context of neuroscience, the physical hardware can be defined as a particular brain input-output node (brain area). Knowing the transformation function of this brain node would reveal its basic operational principles (its neural computations).

The Volterra series expansion technique for system identification fully captures n^{th} order system dynamics in a relatively agnostic manner (limited assumptions about the system). This agnostic approach prevents the generation of partially complete or misleading models. With sufficiently exciting inputs that tap into the system's operational range and a clear definition of the hardware of the system, a unique solution of a system input-output transfer function can be obtained. Note: the term transfer function traditionally refers to linear systems, however, for convenience, this term is extended here to nonlinear systems. The power of this single mathematical formulation is not merely the generation of a predictive model, nor is it insight into

a subset of the tendencies/operations performed by the system. Rather, this function enables comprehensive characterization of the operational range of the system. Examination of this transfer function in the time and frequency domains reveals the operations that the system performs (attenuation, amplification, harmonic generation, frequency intermodulation etc.) Therefore, the relatively agnostic, complete, and rigorous nature of this approach leads to the discovery of a system's universal mathematical rules.

Summarized in this dissertation are a series of system theory-based neuroscience investigations. A deductive approach is initially employed, but after reaching a boundary point with such approach, an inductive approach is used instead. The general theme to the deductive neuroscience investigations is memory processing (how the brain efficiently stores and organizes information from our diverse experiences, encodes temporal order of events, and discriminates between similar memories during memory formation and retrieval). However, the theme generalizes when an inductive approach is employed, becoming focused on comprehensively characterizing system operations, defining the system as a brain area heavily implicated in memory, irrespective of behavior. Investigations occurred at several scales (behavioral measurements to recording neural electrical activity) and in different species (mouse brain slices, behaving rodents, and in vivo human intracranial recordings).

Chapter 2 discusses the first study whereby a novel behavioral paradigm for mice (and rodents more generally) was invented. The purpose for its invention was to advance the animal enrichment model (enriched life experience) of a hypothetical construct in humans—the cognitive reserve hypothesis—that enriched life experiences (number of hobbies, bilinguality, integration in social communities, etc.) serves a neuroprotective role against pathology and symptomology of debilitating diseases including dementia. However, there is a paucity in translating rodent enrichment findings to clinical settings. This is possibly due to two reasons 1) inability to quantify degree of animal engagement with enrichment stimuli using traditional enrichment methods (group housing and larger cages with toys and deeper bedding), 2) rodent

enrichment studies focused on the effects of enrichment on lower order cognition (recognizing a novel object, recalling the location of a platform, etc.). We addressed these two gaps by enriching animals through a new paradigm, the Obstacle Course (OC), which enables controlled enrichment, and by testing animals on tasks of higher order cognition after the OC experience. The Obstacle Course is a square track with built-in easily exchangeable obstacles which mice must overcome to finish a lap and receive reward. Mice running this course 1 hour daily for ~2 months outperformed traditionally enriched mice on behavioral tasks tested, and the OC paradigm revealed novel effects of enrichment on higher order cognition.

A transition occurs in chapter 3, whereby in addition to behavior, neural electrical recordings were acquired and aligned in time to such behavior, and the animal model was rats instead of mice. While memory for spatial navigation has gained much attention in the field, memory for temporal order is relatively less understood. More specifically, the frequency content in the brain has been linked to aspects of memory for spatial navigation. However, its links to memory for temporal order are largely unknown. The goal of this study was to identify the spectral content associated with memory processing for temporal order of events. To test this, electrical activity from the rat brain was recorded while rats behaviorally indicated whether a presented odor is in the right position of an odor-sequence rats had learned prior. Using spectral analysis, we found that 20-40 Hz power was reliably recruited in the hippocampal CA1 area during odor processing. Instantaneous power reached steady-state after a decision was made about temporal order, and power in this post-decision period increased with knowledge of the sequence. Machine learning analysis revealed that power during this post-decision state is predictive of accuracy regarding the temporal order of events. Moreover, 20-40 Hz power was stronger in one end of the CA1 (proximal) axis, which suggests anatomical candidates for its generation. Here, we extended the likely role of 20-40 Hz in gating information within a brain network (suppressing distractors and maintaining representations during sensori-motor integration) to hippocampal sequence memory processing. Its presence in the hippocampus

may have implications for memory impairments related to inattention or inability to suppress distractors. What remains unclear is the necessity for this signal in sequence memory judgements and the information it may be preserving.

Chapter 4 further advances in the animal model to human subjects and focused on a different aspect of memory processing, namely pattern separation. Pattern separation is thought to be a mnemonic process whereby highly similar memories are encoded, stored, and retrieved as distinct experiences despite their high degree of overlap. This is thought to be a core aspect of memory processing, since our day-to-day experiences are highly similar in nature (ex. seeing the same co-workers, at approximately the same time of day and in the same location). Prior studies identified anatomically some of the key brain areas involved in this process, however, what remained unclear were the neural population level firing patterns in these nodes and more generally, in the two key memory systems (hippocampus and neocortex), that facilitate pattern separation. We started by identifying population level synchronous firing patterns in candidate areas (hippocampus and neocortex) that support pattern separation. This was done by recording electrical activity from humans who had undergone a surgical procedure for electrode insertion in the brain for clinical purposes (epilepsy foci identification), while individuals performed a pattern separation task. Using spectral analysis, we found that 4-5 Hz power is increased in both the hippocampus and the neocortex, the two presumed key memory systems, specifically when subjects discriminated between similar items (pattern separation). Using information theory and the 4-5 Hz phase signal, we found that there is an ongoing exchange in communication between the two areas during both encoding and retrieval. However, the exchange had biases for each phase. During memory encoding, the directional bias favored the hippocampus → neocortex direction of information transfer, while the converse was observed in retrieval. This suggests that the two systems are constantly exchanging information, and it is possible that during encoding, information is integrated into neocortex for storage and during retrieval it is accessed from the neocortex by the hippocampus for a comparative measure of

whether items are similar or different. The neural synchrony identified, and the information transfer at this neural synchrony reveal some system dynamics that occur during pattern separation. What remains to be determined is whether these dynamics are necessary for pattern separation to occur, and more generally, the entirety of the neocortical and hippocampal memory systems operations that support this mnemonic process and others.

The investigation in chapter 4 revealed that the two key memory systems, hippocampus and neocortex, both display synchronous 4-5 Hz firing during encoding and retrieval for items that are successfully discriminated between their similar counterparts. This prompted the question of whether the 4-5 Hz in each system occurred independently or if communication between them took place through this signal. As mentioned above, an information theory-based analysis method was employed to make inferences about directional communication. Chapter 5 focuses more closely on this topic of estimating the directional influence in the brain; discussed are two novel adaptations of two existing methodologies for measuring communication between brain areas. First, the information-theory based phase transfer entropy metric was adapted into a time-dependent one. Second, the general linear Kalman filter (which could handle multi-channel and multi-trial data present in the study design) was used to estimate the parameters of an autoregressive model and the model was then used in the granger causality calculation for estimates of directional influence.

The investigations summarized in chapters 2-6 were a key steppingstone for developing an intuition of the unknowns about brain memory systems and considerations regarding how to efficiently uncover these unknowns. Tremendous progress in the field and through this prior work has been made. Yet we are far from a full characterization of these systems and many other systems in the brain, such that we can build them into physically realizable circuits. Therefore, a different approach is taken next. First, the identification of the majority of computations in a brain area are first investigated, in a relatively complete manner and with as

little assumptions as possible. Then, ultimately, the characterized system can teach the researcher about its operational capacities and their behavioral consequences.

In chapter 6, I discuss the use of system identification to uncover as much of the operational dynamics as possible of a candidate brain area, the hippocampal CA1. This requires a clear definition of the system hardware (brain area) and the input and output data (causal link); therefore, a transition is made in the animal model, going from humans to live slices of mouse brains (slice electrophysiology). Volterra series expansion technique is employed on input-output data recorded from mouse CA1 brain slices. Input times were drawn from a Poisson distribution and uniformly varied in amplitude between two values and were delivered via a stimulating electrode through the CA1's main input pathway (hippocampal CA3 axonal projections). Induced outputs from the CA1 (apical dendritic field) were recorded. Despite the CA3:CA1 node being arguably the most heavily studied synapse in the brain, we made discoveries about the CA1, through this system identification study, never before known in neurophysiology.

Contrary to popular believe, the CA1 is a nonlinear system. Examination of its second order Volterra kernel revealed system properties, namely how it treats the inputs based on its content (amplify, suppress, frequency filtering, etc.). In addition to the insight gained, the Volterra series-based system identification enabled the capturing of most system dynamics in a single mathematical function. This function can predict with 93% accuracy the CA1 output waveform (instantaneous values) to random inputs delivered in novel animals.

The dynamics of CA3:CA1 node are a result of immense biological complexity – more complicated than the most complicated machine. The ability to condense such complexity in the form of Volterra kernels was an open question in the field of neurophysiology. Having successfully done this in one node provides a steppingstone for a lifetime of work, as discussed in the concluding chapter. It opens the door to the possibility of repeating the procedure in the remaining nodes. A possible direction this work may take is to identify two key transfer functions

– one at the most fundamental scale, and the second at the most functional scale. The transfer function at the fundamental scale is one that maps the basic unit of communication in the brain – input spike patterns to output spike patterns between a pair of neurons. The transfer function at the functional scale is a net input-output transformation function that uncovers how inputs from the external world are processed into outputs that support our subsequent interaction with the world we live in. Knowing these, we can possibly achieve a more full understanding of the brain and relieve suffering from debilitating conditions of neurology and neuropsychiatry.

Chapter 2: Using behavior to understand rodent behavior (deductive):

A new enrichment paradigm reveals higher order cognitive enhancements and enables controlled delivery of enrichment - an advance to the rodent enrichment model

2.1 Abstract

Enrichment in rodents affects brain structure, improves behavioral performance, and is neuroprotective. Similarly, in humans, according to the cognitive reserve concept, enriched experience is functionally protective against neuropathology. Despite this parallel, the ability to translate rodent studies to human clinical situations is limited. This limitation is likely due to the simple cognitive processes probed in rodent studies and the inability to control, with existing methods, the degree of rodent engagement with enrichment material. We overcome these two difficulties with novel behavioral tasks that probe, in a fine-grained manner, aspects of higher order cognition associated with deterioration with aging and dementia, and a new enrichment protocol, the 'Obstacle Course' (OC), which enables controlled enrichment delivery, respectively. Together, these two advancements will enable better specification (and comparisons) of the nature of impairments in animal models of complex mental disorders and the potential for remediation from various types of intervention (e.g., enrichment, drugs). We found that two months of OC enrichment produced substantial and sustained enhancements in categorization memory, perceptual object invariance, and cross-modal sensory integration in mice. We also tested mice on behavioral tasks previously shown to benefit from traditional enrichment: spontaneous object recognition, object location memory, and pairwise visual discrimination. OC enrichment improved performance relative to standard housing on all six tasks and was in most cases superior to conventional home cage enrichment and exercise track groups.

2.2 Introduction

Decades of research findings support the conclusion that experience affects neural architecture and plasticity [1-5]. Indeed, enriched experiences have been shown to affect brain morphology and electrophysiology, promote positive effects on behavior and can be protective against the effects of aging, Alzheimer's disease (AD), other forms of dementia, and a myriad of additional disease states [6-14]. Although predominantly studied in animal models, similar findings have been reported in humans [15-17], where factors including education level, vocabulary, hobbies and multi-linguality serve as composite measures of 'cognitive reserve' – an operational construct that has been posited to play a protective role against aging and dementia [18].

Despite the overwhelming evidence regarding the beneficial effects of enrichment from rodent studies, and the cognitive reserve hypothesis pointing to promising protective effects in humans, there is limited translatability to human clinical situations. There are two aspects to the rodent environmental enrichment (EE) model that likely contribute to its limited clinical translatability.

First, most rodent studies have investigated enrichment effects on lower-order cognitive processes using multiple variants of the same simple behavioral tasks. These behavioral tasks primarily include finding a platform in a water maze (Morris Water Maze; MWM, probing spatial learning and memory), open field exploration (open field test; OFT, probing exploration and emotionality), elevated mazes (elevated plus maze; EPM, probing anxiety) and object recognition (OR, probing recognition memory) [19]. The output of this body of work is impressive and has established that EE improves spatial and non-spatial learning and memory, sensory discrimination, sociability, and anxiety, in wild type rodents and/or models of aging and disease [6, 19-26]. However, there is no fine-grained analysis of the effects of EE on complex aspects of cognition. This is needed for a better behavioral mapping between rodents and humans and will

more clearly point to which aspects of complex mental disorders benefit from EE (or other forms of therapy). Here, we employ three new behavioral tasks probing different aspects of higher order cognitive processes which are subject to deterioration due to aging and AD in humans [27-29]: 1) categorization and abstraction -- hypothesized as the functional basis for the neocortical-hippocampal memory network in memory formation, consolidation and retrieval ; 2) multisensory integration, which requires inter-regional communication, a prerequisite for engagement of the aforementioned network; and 3) higher order perception.

Second, with existing methods of enrichment, it is not possible to control for the degree of animal engagement with enrichment stimuli or identify which aspects of enrichment contribute to the measured outcomes. It is important to have a method which controls for animal engagement with objects to study how factors such as age, disease or brain manipulations interact with EE effects, because such variables can influence animals' motivation level and therefore, receipt of enrichment. Moreover, despite recent improvements in standardization of enrichment methods [30, 31], as opposed to traditional enrichment (placing animals in larger cages, group housed, with extra bedding, toys, and running wheels), there is not an existing method that enables parsing out social stimulation, cognitive stimulation and physical exercise. Here, we present a novel enrichment paradigm, the Obstacle Course, which addresses these two foregoing limitations to existing enrichment methods.

We advance the EE model's utility in two ways: 1) showing that enrichment also enhances specific aspects of higher order cognitive functions in mice by using a series of complex behavioral tasks, and 2) delivering enrichment through a novel paradigm, the Obstacle Course (OC), which enables systematic control over animal engagement with enrichment stimuli, is delivered outside of animal housing, and has an exercise matched Control Course (CC) (Fig. 1). We tested enrichment effects on categorization, perceptual invariance, and multisensory integration using new paradigms for object category, view-invariance, and crossmodal object recognition, respectively (Table 1). We found that OC training elicited

improved and long-lasting performance on the three aforementioned tasks, especially at longer delays between encoding and retrieval. These effects were not observed with exercise alone, tested using the CC. Lastly, because these enrichment effects were observed with our novel OC enrichment paradigm, we also validated the OC by replicating findings obtained from three previously utilized tasks: object recognition, object location memory, and pairwise visual discrimination. For all tasks, enrichment was delivered both through the OC and through traditional enriched housing (EH). The OC group was comparable to the EH group on tasks with shorter learning-to-testing delays and outperformed the EH group on most tasks of longer delays.

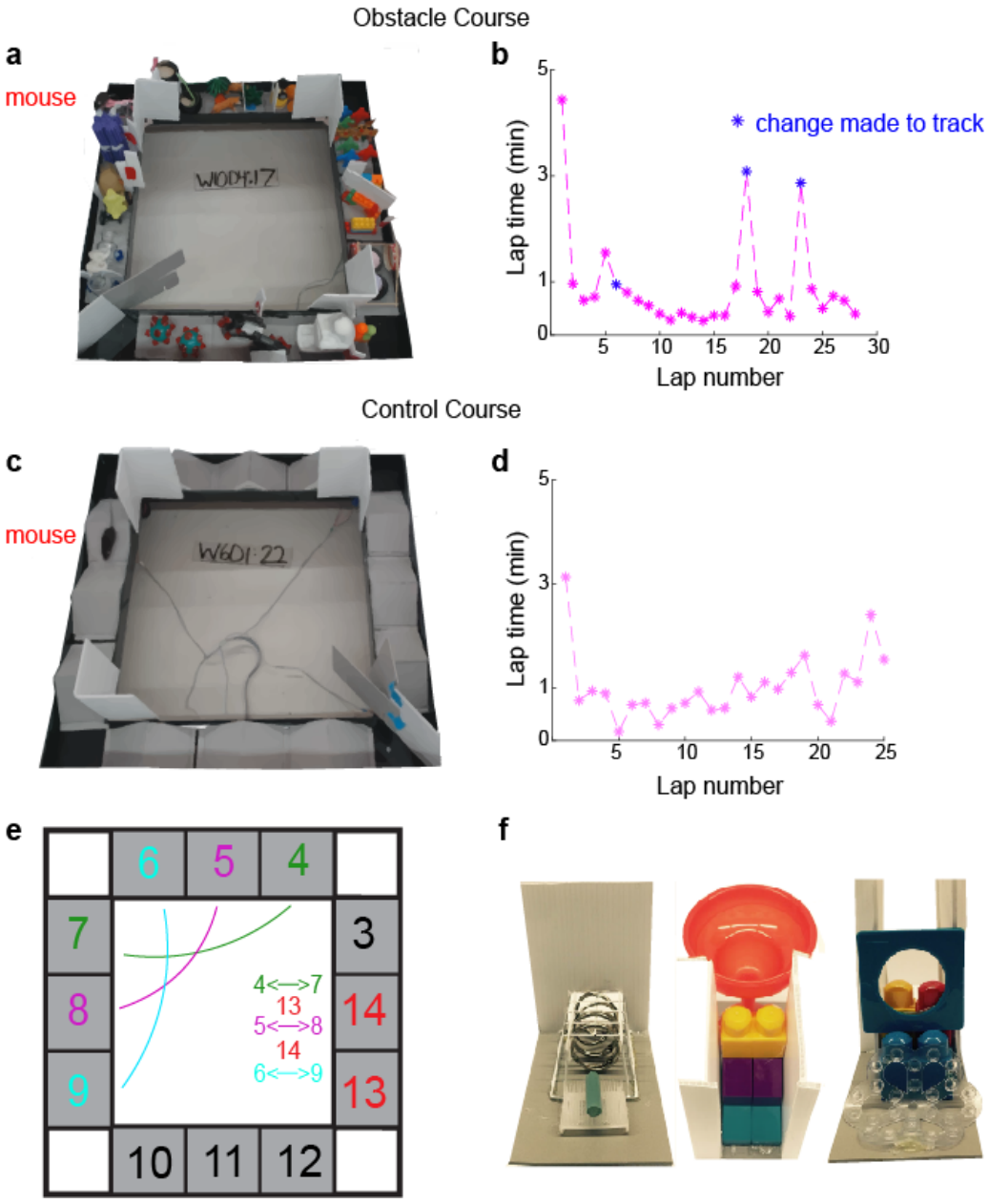


Figure 1. The obstacle course paradigm
a. Example OC track filled with 12 inserts. **b.** Example lap time data for a 30-min duration on the OC. Note that animals tended to take longer to complete a given lap when a change to the track was performed immediately prior to the lap (blue asterisk). See Supplemental Video 2 for demonstration of animal change in speed pre and post change to the track. **c.** An example CC track filled with 12 hurdles. **d.** Example lap time data for a 30-min duration on the CC. No changes were made to the CC throughout a given session**. e.** Schematic of changes made in a 1-h OC session. A total of 5 changes were made, which included 3 rearrangements (e.g., swapping obstacle 4 with obstacle 7, indicated by arrows) interleaved with 2 replacements of old inserts with novel ones (indicated in red). See Supplementary Figure 1 for a detailed OC and CC protocol. **f.** Three example obstacles used in the OC (12 for the entire course).

Event	Date	Age of mice	Duration since discontinuation of enrichment
First day of enrichment	10/08/18	1.5 m	-
Last day of enrichment	12/15/18	4 m	-
First day of SOR testing during EE	12/3/18	4 m	0
Last day of SOR testing during EE	12/15/18	4 m	0
SOR rep post-EE	1/12/19	5 m	28 d [4 wks]
SOR rep post-EE	1/15/19	5 m	31 d [4.4 wks]
First day of OCR testing	02/01	6 m	48 d [6.8 wks]
Last day of OCR testing	03/01/19	6 m	75 d [10.7 d]
First day of VIOR testing	03/06/19	7 m	81 d [11.6 wks]
Last day of VIOR testing	03/20/19	7 m	95 d [13.6 wks]
First day of CMOR testing	04/10/19	8 m	116 d [16.6 wks]
Last day of CMOR testing	04/23/19	8 m	129 d [18.4 wks]
First day of OL testing	05/06/19	9 m	142 d [20.3 wks]
Last day of OL testing	06/16/19	10 m	183 d [26.1 wks]
First day of PD testing	06/24/19	10 m	191 d [27.3 wks]
Last day of PD testing	8/1 - 10/19	14 - 16 m	231 d [33 wks]

Table 1. Experimental timeline
Table outlining enrichment and behavioral testing schedule, including experiment date, age of mice during each experiment, and duration between enrichment discontinuation and behavioral testing. Abbreviations: rep-repeated, m-months, wks-weeks, d-days.

2.3 Results

Spontaneous object recognition (SOR): Enhancement of long-term object memory with OC training

We first tested whether the OC experience can induce the previously reported EE enhancements of object recognition memory [32, 33]. Four groups of mice, Obstacle Course (OC), Control Course (CC), Enriched housing (EH), Standard housing (SH), were first tested on the SOR task with a 10-min sample phase and a 24-h delay in the 9th week of enrichment (Fig. 2a-b). In this first experiment, an analysis of variance (ANOVA) showed that the effect of group did not reach significance, $F(3, 36) = 2.613$, $p = 0.066$, partial $\eta^2 = 0.179$. However, planned

comparisons using independent samples t-tests indicated that the OC group showed an enhancement on task performance compared to the SH group, $t(18) = -3.027$, $p = 0.007$, after Bonferroni correction for multiple comparisons ($p = 0.0083$). Two-tailed paired-samples t-tests (sample DR vs. choice DR) showed that mice in the OC, $t(9) = -8.109$, $p < 0.001$, EH, $t(9) = -3.624$, $p = 0.006$, and SH groups, $t(9) = -3.846$, $p = 0.004$, all demonstrated evidence of intact memory.

In a subsequent experiment, we introduced both short and long-term retention delays and increased task difficulty by shortening the sample phase to 7 min (Fig. 2a, c). A repeated measures ANOVA revealed a significant interaction between length of delay and group, $F(3, 36) = 16.818$, $p < 0.001$, partial $\eta^2 = 0.584$. Significant main effects were also found for delay, $F(1, 36) = 124.724$, $p < 0.001$, partial $\eta^2 = 0.776$, and group, $F(1, 36) = 5.720$, $p = 0.003$, partial $\eta^2 = 0.323$. With the 20-min delay, no significant differences between groups were found and a two-tailed paired samples t-test showed that the OC, $t(9) = -7.364$, $p < 0.001$, CC, $t(9) = -4.784$ $p = 0.001$, EH, $t(9) = -11.927$, $p < 0.001$, and SH groups, $t(9) = -3.114$, $p = 0.012$, all demonstrated intact object memory. However, between-group differences were observed with the 24-h delay. Further analysis with *post-hoc* independent samples t-tests revealed significant differences in performance between the OC group and the EH group, $t(18) = -6.287$, $p < 0.001$, CC group, $t(18) = -5.573$, $p < 0.001$ and SH group, $t(18) = -5.836$, $p < 0.001$, when the delay was 24 h. Two-tailed paired-samples t-tests showed that only the OC group discriminated the novel object $t(9) = -8.075$, $p < 0.001$.

One month following the end of the enrichment period, the enhancement observed on the SOR task with a 7-min sample and 24-h delay persisted for the OC group (Fig. 2a, d). Observed were significant main effects of group, $F(3, 36) = 3.575$, $p = 0.023$, partial $\eta^2 = 0.230$, and delay, $F(1, 36) = 7.421$, $p = 0.010$, partial $\eta^2 = 0.171$, while the interaction between group and delay did not quite reach significance, $F(3, 36) = 2.745$, $p = 0.057$, partial $\eta^2 = 0.186$. Further analysis with planned independent samples t-tests revealed a significant difference in

performance on the task between the OC group and the EH group, $t(18) = -5.292$, $p < 0.001$ and the CC group, $t(18) = 3.917$, $p = 0.001$ when the delay was 24 h. At this delay, only the OC group performed significantly above chance, $t(9) = -5.962$, $p < 0.001$. When the delay was reduced to 20 min, each group's performance, OC, $t(9) = -2.750$, $p = 0.022$, CC, $t(9) = -4.397$, $p = 0.002$, EH, $t(9) = -2.773$, $p = 0.022$, and SH, $t(9) = -4.054$, $p = 0.003$, was above chance on the SOR task.

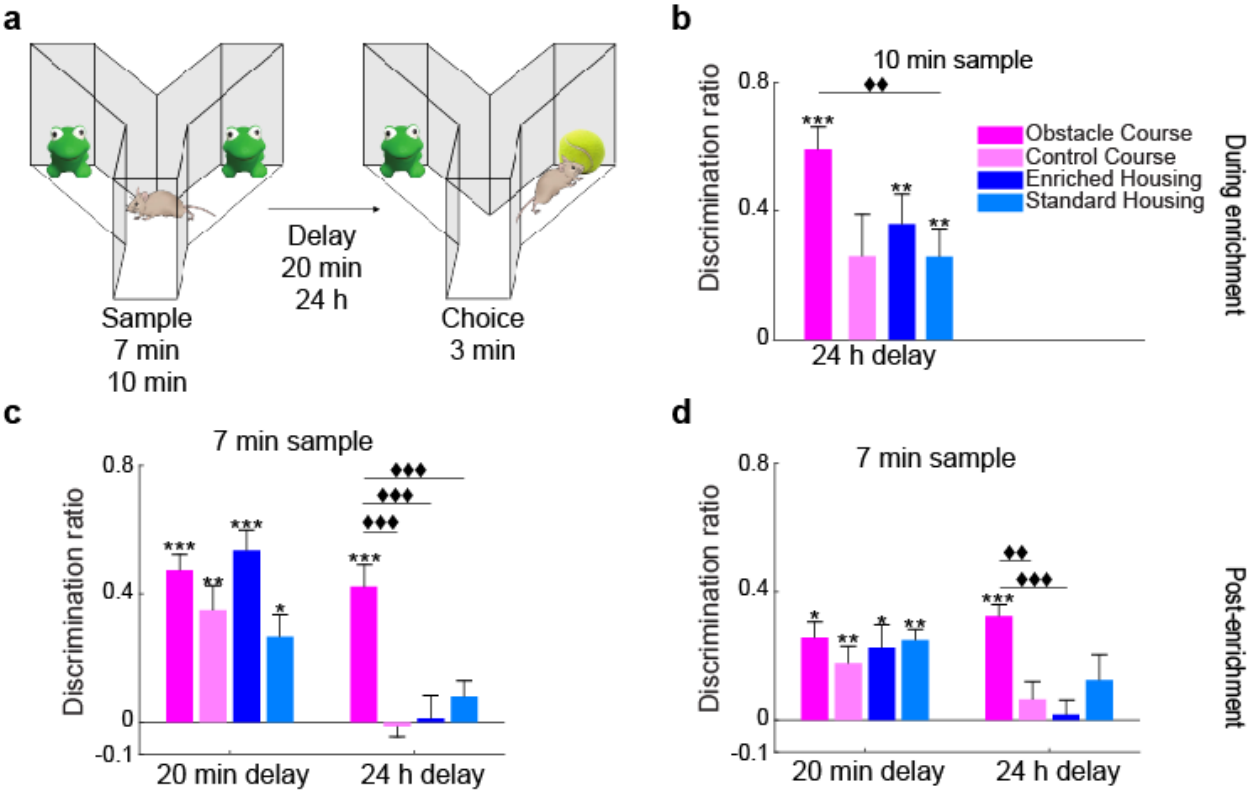


Figure 2. SOR - Enhancement of long-term object memory with OC training exceeds EH effects

a. Schematic representation of the Spontaneous Object Recognition (SOR) experiments. **b.** Task performance with a 10-min sample and a 24-h delay during the 9th week of enrichment. The OC group demonstrated a significantly higher discrimination ratio (DR) compared to the SH group. OC, EH, and SH groups showed DRs above chance. **c.** SOR task performance with a sub-optimal sample time at two delay durations during the 10th week of enrichment. With a 20-min delay, all groups significantly discriminated above chance. With an increased delay to 24 h, the OC group DR was significantly higher compared to all other groups. (Delay x group interaction $p<0.001$, delay main effect $p<0.001$, and group main effect $p = 0.003$). Only the OC group showed a DR above chance. **d.** Replication of c, but completed 1 month post-discontinuation of enrichment. With a 20-min delay, all groups significantly discriminated above chance, demonstrating intact object memory. With an increased delay to 24 h, the OC group demonstrated enhanced object memory, performing significantly better than the EH and CC

groups (Delay x group interaction $p = 0.057$, delay main effect $p = 0.01$, and group main effect $p = 0.023$). The OC group was the only group to discriminate the novel object significantly above chance.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ indicate significant two-tailed paired-sample t-tests (choice DR vs. sample DR). ♦♦ $p < 0.01$, ♦♦♦ $p < 0.001$ indicate a significant difference between groups (two-tailed independent samples t-test, Bonferroni correction $p = 0.0083$).

Object location memory (OLM): Enhanced object spatial memory with OC training

For the second validation experiment, we investigated whether OC training can elicit enhancements in spatial memory as previously demonstrated with EE [34], and more specifically, object location enhancements [32]. To do so, all groups performed the OLM task (Fig. 3a). Enhanced task performance after OC training was observed when the delay between sample and choice was increased from 20 min to 24 h (Fig. 3b). With the same sample time of 10 min, a repeated measures ANOVA revealed a significant interaction between the length of delay and group, $F(3, 36) = 4.294$, $p = 0.011$, partial $\eta^2 = 0.264$, and a significant main effect of delay, $F(1, 36) = 30.763$, $p < 0.001$, partial $\eta^2 = 0.461$; however, no main effect of group, $F(3, 36) = 0.963$, $p = 0.421$, partial $\eta^2 = 0.074$ was found (Fig 3d.*i*).

With a 20-min delay, no significant differences between groups were observed (Fig. 3d.*i*). Two-tailed paired-samples t-tests (sample DR vs. choice DR) showed that the CC, $t(9) = -5.279$, $p = 0.001$, EH, $t(9) = -5.394$, $p < 0.001$, and SH groups, $t(9) = -6.056$, $p < 0.001$, all discriminated the displaced objects, while discrimination by the OC, $t(9) = -2.224$, $p = 0.053$, did not quite reach significance. Subsequent planned analysis for the 24-h delay using independent samples t-tests revealed significant differences in performance on the task between the OC and CC groups, $t(18) = 4.340$, $p < 0.001$, as well as the EH and CC groups, $t(18) = 3.136$, $p = 0.006$. Moreover, with this longer delay of 24-h, the enriched groups, OC, $t(9) = 5.092$, $p = 0.001$ and EH, $t(9) = 4.464$, $p= 0.002$, and the SH group, $t(9) = 2.746$, $p = 0.023$, discriminated the displaced object.

Lastly, to probe the differences in spatial memory between the two enrichment groups further, we tested all groups with a shortened sample time of 7 min and delay duration of 24 h.

In this experiment, an ANOVA showed a significant effect of group, $F(3, 36) = 3.623$, $p = 0.022$, partial $\eta^2 = 0.232$, and independent samples t-tests revealed that the OC group showed an enhancement on task performance compared to the EH group, $t(18) = -3.041$, $p = 0.007$ (Fig. 3c). At this sub-optimal sample and increased delay, only mice in the OC group, $t(9) = 2.585$, $p = 0.029$ discriminated the displaced object.

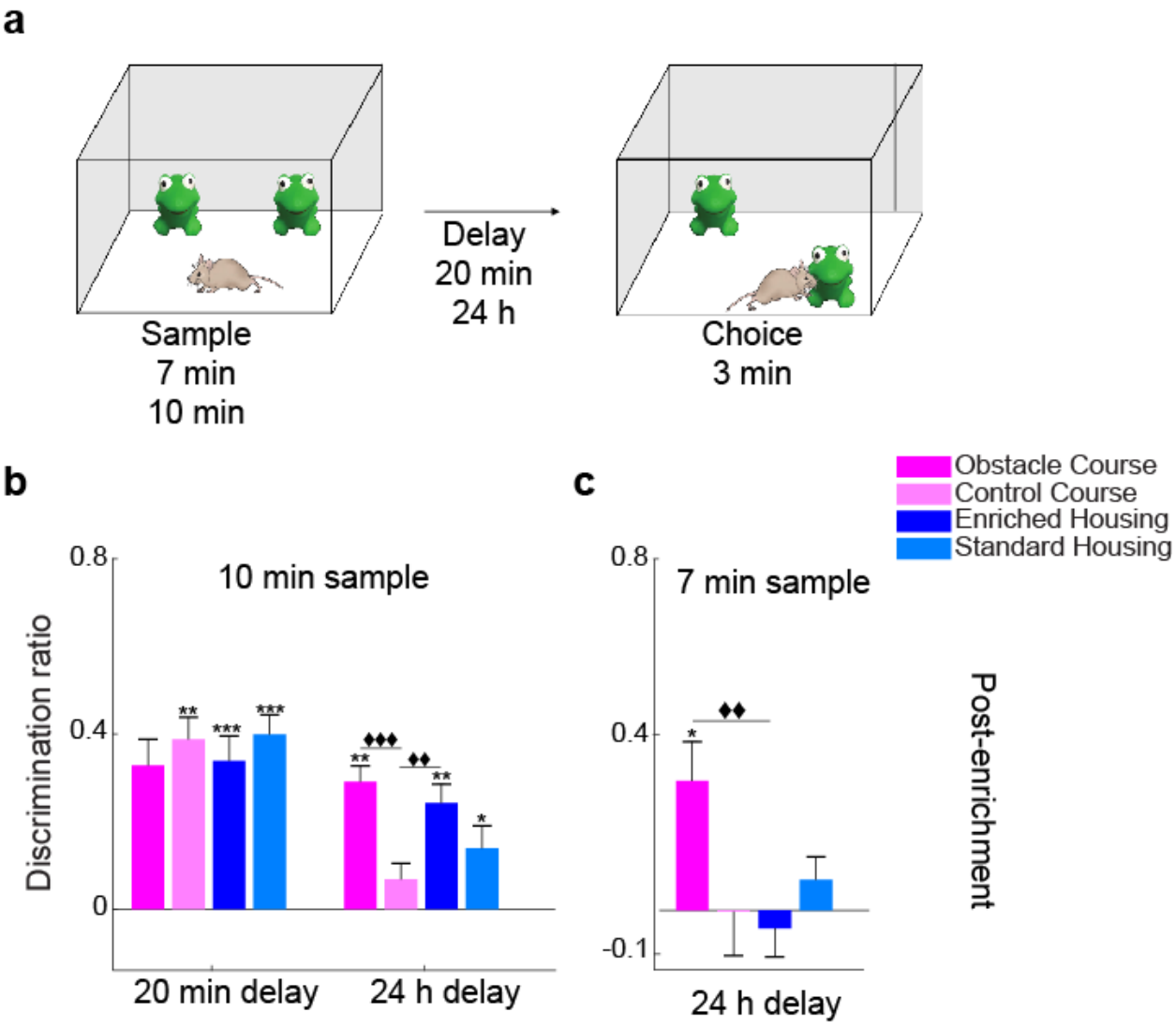


Figure 3. OLM: Enhanced memory for location with OC training exceeds EH effects
a. Schematic of the Object Location Memory (OLM) experiments. **b.** OLM task performance. With a 20-min delay, the CC, EH, and SH groups demonstrated intact spatial memory. With a 24-h delay, the OC and EH groups demonstrated enhanced object location memory, performing significantly better than the CC group (Delay x group interaction $p = 0.011$, delay main effect $p < 0.001$ and group main effect $p = 0.421$). The OC, EH, and SH groups discriminated the displaced object above chance. **c.** Performance with a 7-min sample and 24 h delay. The OC

group showed enhanced memory relative to the EH group (Group main effect $p = 0.022$). Only the OC group discriminated significantly above chance. Experiment was implemented 5-6.5 months post enrichment.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ indicate significant two-tailed paired-sample t-tests (choice DR vs. sample DR). ♦♦ $p < 0.01$, ♦♦♦ $p < 0.001$ indicate a significant difference between groups (two-tailed independent samples t-test, Bonferroni correction $p = 0.0083$).

Pairwise visual discrimination (PD): Enhanced cognitive flexibility with OC and CC training

For the last replication experiment, we validated whether the OC experience can induce reported EE effects outside the domain of object-based tasks. It was previously shown that the PD task is effective in assessing cognitive function between groups (i.e. sex, disease models [35]), and that EE enhances reversal learning in this task [36]. No significant group differences were observed in the number of pretraining days $F(3,36) = 1.928$, $p = 0.142$, partial $\eta^2 = 0.138$ (habituation stages 1-7), nor in the acquisition phase (in regard to overall accuracy, number of sessions to criterion and session accuracy) (Supplementary Figure 2). With regard to the reversal phase (Fig. 4a.ii), there was a group main effect in the number of sessions to reach criterion (achieving 80% accuracy or higher on two successive sessions), $F(3,36) = 12.819$, $p < 0.001$, partial $\eta^2 = 0.517$ (Fig. 4b), and independent samples t-tests revealed significant differences between the OC and SH groups, $t(18) = -6.104$, $p < 0.001$ as well as the CC and SH groups $t(18) = -5.273$, $p < 0.001$, whereby the OC and CC groups required significantly fewer sessions. Additionally, there was a main effect of session when considering mean group accuracy in each of the first five sessions (where the first mouse reached criterion), $F(3,36) = 61.723$, $p < 0.001$, partial $\eta^2 = 0.632$ (Fig. 4c.*i*), a main effect of group $F(3,36) = 3.209$, $p = 0.034$, partial $\eta^2 = 0.211$, but the interaction between group and session was not significant $F(3,36) = 1.045$, $p = 0.411$, partial $\eta^2 = 0.080$. Post-hoc analysis of session performance revealed an effect of group during session 5, $F(3,36) = 5.606$, $p = 0.003$, partial $\eta^2 = 0.318$. Independent sample t-tests revealed that the OC group performed with significantly higher accuracy compared to the SH group on session five, $t(18) = 3.719$, $p = 0.002$ (Fig. 4.c.*i*). Lastly, there was

an effect of group when considering mean accuracy across the first five sessions, $F(3,36) = 3.222$, $p = 0.034$, partial $\eta^2 = 0.212$ (Fig. 4c.*ii*).

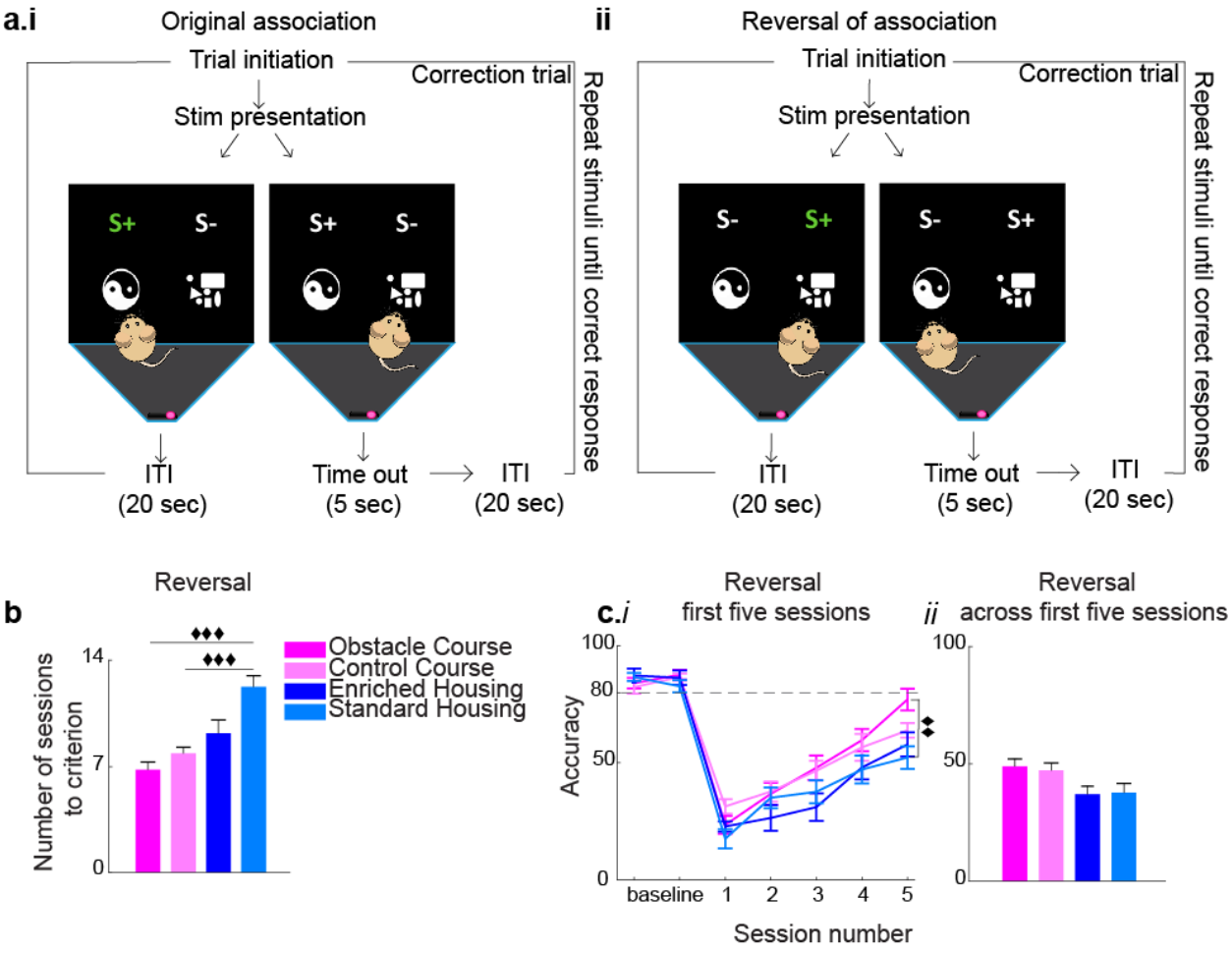


Figure 4. PD task – enhanced cognitive flexibility with OC and CC training exceeds EH effects

a. Schematic representation of the Pairwise visual Discrimination (PD) experiment. **a.i-ii.** Stimulus-reward association learning. **a.i.** Original association: animal receives reward upon contact with the S+ stimulus. **a.ii.** Reversal of association: previously learned S- stimuli become S+ stimuli and associated with reward. **b.** Mean session number to reach criterion (performance of 80% accuracy on two successive sessions) in the reversal phase per group (group effect $p < 0.001$), with the OC ($p < 0.001$) and CC ($p < 0.001$) groups requiring significantly fewer sessions compared to the SH group. **c.i.** Group percent accuracy on the first five reversal sessions, with a main effect of session ($p = 0.003$) and an effect of group on the fifth session ($p < 0.001$), with the OC group performing significantly higher than the SH group on session five. **c.ii.** Mean accuracy across the first five reversal sessions for each group, with a group main effect ($p = 0.034$). Post-hoc independent samples t-test did not reveal group pairwise differences. ♦♦ $p < 0.01$, ♦♦♦ $p < 0.001$ indicate a significant difference between groups (two-tailed independent samples t-test, Bonferroni correction $p = 0.0083$).

Object category recognition (OCR): Enhancement of object category recognition with OC training

Next, we examined OC enrichment effects on arguably more complex cognitive domains than those previously explored. First, we tested whether the OC enrichment enhances object category recognition memory (Fig. 5a) [37]. All groups were tested on two retention delays (30 min, 1 h), both with a 10-min sample. A repeated measures ANOVA revealed a significant interaction between length of delay and group, $F(3, 36) = 7.023$, $p = 0.001$, partial $\eta^2 = 0.369$, and a significant main effect for delay, $F(1, 36) = 14.594$, $p = 0.001$, partial $\eta^2 = 0.288$, but not for group, $F(1, 36) = 0.745$, $p = 0.532$, partial $\eta^2 = 0.058$ (Fig. 5b). With the 30-min retention delay, two-tailed paired-samples t-tests (sample DR vs. choice DR) showed that mice in the CC, $t(9) = -6.135$, $p < 0.001$ EH, $t(9) = -3.129$, $p = 0.012$, and SH groups, $t(9) = -6.074$, $p < 0.001$ all demonstrated intact object category memory, while the OC, $t(9) = -1.876$, $p = 0.093$, performance did not reach significance. However, between-group differences were observed when the retention delay was increased to 1 h (Fig 5d). Further analysis with planned comparisons using independent samples t-tests revealed a significant difference in group performance on the task between the OC and the SH group, $t(18) = 3.884$, $p = 0.001$, when the delay was 1 h. Two-tailed paired-samples t-tests showed that the OC $t(9) = -5.895$, $p < 0.001$ and the EH $t(9) = -2.937$, $p = 0.017$, groups discriminated the novel category.

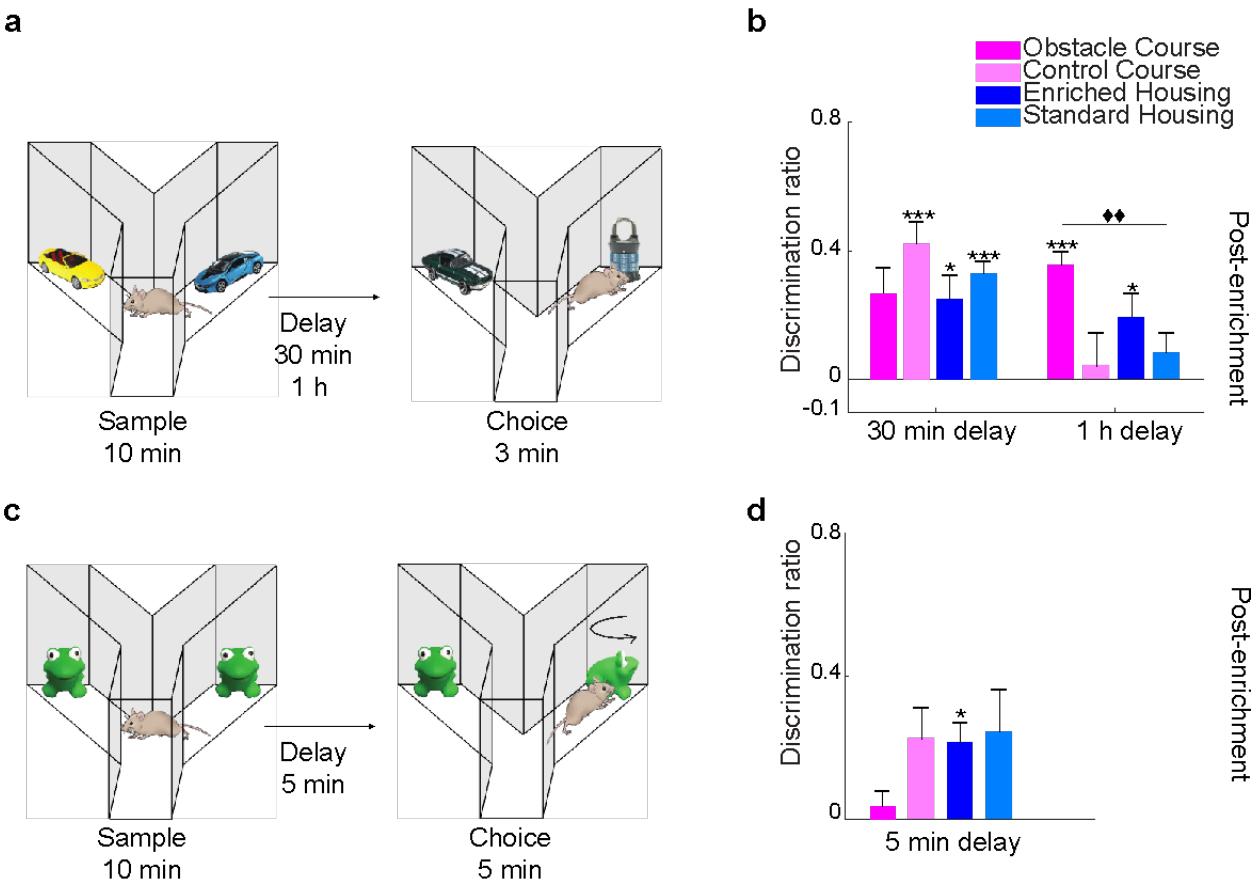


Figure 5. OCR & VIOR - Enhancement of both object category recognition (OCR) and object recognition despite rotational changes to objects (VIOR) with OC training
a. Schematic representation of the Object Category Recognition (OCR) experiment. **b**. OCR task performance with two delay durations. With a 30-min delay, the CC, EH, and SH groups significantly discriminated above chance, demonstrating intact object category memory. With an increased delay to 1 h, the OC group demonstrated enhanced object category memory, performing significantly higher compared to the SH group (Delay x group interaction $p = 0.001$, delay main effect $p = 0.001$, and group main effect $p = 0.532$). The OC and EH groups significantly discriminated the novel category above chance. OCR experiments were implemented 1.5-2.5 months post-discontinuation of enrichment. **c.** Schematic representation of the View Invariant Object Recognition (VIOR) experiment. **d**. VIOR task performance. The OC group treated the rotated object as familiar, equally exploring rotated and unrotated objects in the choice phase. The EH group significantly discriminated the rotated object. No significant group effect was observed ($p = 0.240$). VIOR experiments were implemented 3 months post-discontinuation of enrichment.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ indicate significant two-tailed paired-sample t-tests (choice DR vs. sample DR). ♦♦ $p < 0.01$, ♦♦♦ $p < 0.001$ indicate a significant difference between groups (two-tailed independent samples t-test, Bonferroni correction $p = 0.0083$).

View-invariant object recognition (VIOR): Enhancement of object recognition despite rotational changes in object orientation

We then investigated the effects of OC enrichment on higher order perception using VIOR testing (Fig. 5c). An ANOVA showed that the effect of group did not reach significance, $F(3, 36) = 1.468$, $p = 0.240$, partial $\eta^2 = 0.109$. The OC group, to the most extent (numerically), treated both rotated and non-rotated objects as familiar, exploring them equally (Fig. 5d). Conversely, a paired-samples t-test showed that the EH group displayed preference for the rotated object suggesting that this group tended to perceive the rotated object as novel $t(9) = -3.046$, $p = 0.014$.

Cross-modal object recognition (CMOR): Enhanced cross-modal object memory with OC training

Lastly, we identified the degree to which OC enrichment facilitates integration of sensory information from different modalities by testing all groups on a tactile-to-visual CMOR task (Fig. 6a). Enhanced task performance after OC training was observed when the delay between sample and choice was increased to 1 h (Fig. 6b). A repeated measures ANOVA did not show a significant main effect of delay, $F(1, 36) = 1.584$, $p = 0.216$, partial $\eta^2 = 0.042$, and the interaction between length of delay and group did not reach significance $F(3, 36) = 2.762$, $p = 0.056$, partial $\eta^2 = 0.187$. However, there was a significant main effect of group $F(3, 36) = 4.981$, $p = 0.005$, partial $\eta^2 = 0.293$.

At a 30-min delay, planned independent samples t-tests revealed a significant difference in group performance on the task between the CC and the SH group, $t(18) = -3.758$, $p = 0.001$ (Fig. 6b). At the same delay, two-tailed paired-samples t-tests showed that both the OC, $t(9) = -3.720$, $p = 0.005$ and the SH, $t(9) = -4.468$, $p = 0.002$ groups discriminated between the novel and familiar objects. Subsequent planned analysis for the 1-h delay using independent samples t-tests revealed a significant difference in group performance on the task between the OC and the CC group, $t(18) = 4.119$, $p = 0.001$. Moreover, at this 1-h delay, only the OC group discriminated between the novel and familiar objects $t(9) = -4.357$, $p = 0.002$.

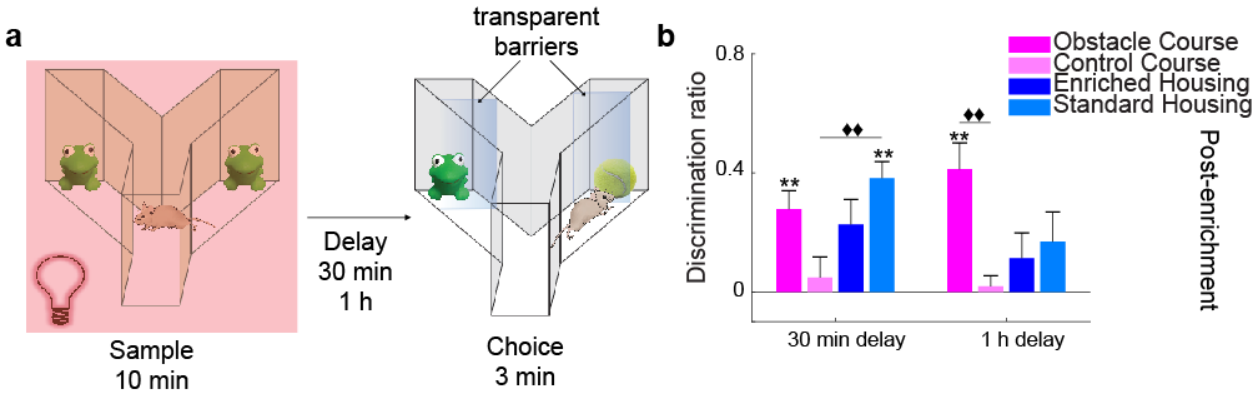


Figure 6. CMOR: Enhanced cross-modal object memory with OC training
a. Schematic of the Cross-Modal Object Recognition (CMOR) experiment. **b**. CMOR task performance. At a 30-min delay, the SH group discriminated significantly more than the CC group. The OC and SH groups discriminated the novel object above chance. With a 1-h delay, the OC group demonstrated enhanced performance relative to the CC group (Delay x group interaction $p = 0.056$, delay main effect $p = 0.216$, and group main effect $p = 0.005$). Only the OC group could discriminate the novel object. Experiment was implemented 4-4.5 months post enrichment.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ indicate significant two-tailed paired-sample t-tests (choice DR vs. sample DR). ♦♦ $p < 0.01$, ♦♦♦ $p < 0.001$ indicate a significant difference between groups (two-tailed independent samples t-test, Bonferroni correction $p = 0.0083$).

2.4 Discussion

In the present study, we sought to advance the rodent enrichment model by performing a fine-grained analysis on the effects of EE on different aspects of higher order cognition using the OC as a novel protocol enabling controlled delivery of enrichment. We first showed that the OC course is at least comparable or superior to traditional home cage enrichment in terms of the benefits conferred on performance in established learning and memory tasks. Then we presented a novel account of enrichment-mediated enhancements of lesser studied cognitive functions: object categorization/generalization, object perceptual processing and information transfer across sensory networks. The extension of enrichment effects to these domains should help to improve behavioral mapping between rodents and humans in that they are more complex functions associated with deterioration from aging and dementia [27-29]. This kind of

fine-grained, systematic analysis using novel behavioral tasks will be critical in revealing *which* aspects of complex mental disorders could benefit from enrichment or other forms of therapy. The present study takes important first steps in revealing the potential for beneficial effects of enrichment interventions across a spectrum of complex cognitive abilities. Lastly, delivery of enrichment in a controlled manner, accomplished with the OC protocol, can provide greater insight into which aspects of enrichment and how the degree of engagement with stimuli influences certain processes.

Our results were observed with a novel enrichment paradigm, the Obstacle Course. Conversely, Enriched Housing did not elicit improvements in view-invariant object processing nor cross-modal sensory integration, and yielded a less prominent effect on category recognition. The OC group also outperformed current standards of enrichment on previously reported measures, demonstrating superior long-term memory for object identity and spatial location (especially with short, 'suboptimal' study durations) and enhanced cognitive flexibility (pairwise discrimination reversal). Most of these effects cannot be explained by physical exercise alone, as tested through an exercise-matched Control Course. Lastly, OC induced enhancements were sustained for over 8 months after discontinuation of enrichment, an important demonstration of the potential longevity of effects produced by this protocol.

Across tasks, the main enhancement effects with OC training that significantly differed from EH were generally observed with shorter sample times and longer delays (SOR, OCR, CMOR, and OL). This suggests that, with OC training, less time is necessary to encode and/or consolidate information about the external environment into long-term memory. In line with this, is the observed result of superior performance by the OC group compared to all other groups on OCR. We have previously reported that mice can only perform OCR with retention delays longer than 30 min when an exemplar pre-exposure period is implemented, which possibly enables an improved representation of the to-be-remembered object categories [37]. However, in the current study, the OC group was able to perform the task successfully without an exemplar

preexposure period. Thus, the benefit of enrichment in general, and our OC paradigm in particular, could be explained at least partly by improved schema formation.

Schema formation is a process reliant on semanticization and abstraction of experiences and generation of internal representations which are categorically organized. The facilitation of these processes could enable items to be more rapidly integrated into long-term memory if they are to some degree consistent with existing memories or schemas [38-40]. The OC experience requires rich interactions with a wide variety of object-based obstacles, which may facilitate establishment of object schemas. Thus, a reasonable hypothesis is that OC animals have more highly diversified internal representations of the world that would support such an effect. This possibility, as well as alternatives involving enhanced learning and memory via boosting of neuroplasticity mechanisms, warrant further testing in subsequent studies to unravel the means by which OC training achieves superior effects.

OC training also improved perception and multi-sensory integration. With OC training, mice recognized object identity despite rotation (VIOR) to a greater degree than observed with EH. This finding also supports the notion that OC mice generated a stronger internal representation and/or better retained object representations, recognizing a given object from a different angle. Another possibility is that more sophisticated perceptual processes were at play and not merely superior long-term memory [41, 42]. Additionally, with OC enrichment, mice were better able to perform an explicitly cross-modal cognitive task, visually recognizing objects that they had previously only explored tactilely. The mechanisms by which these enhancements occurred cannot be delineated from the present study. However, these results demonstrate that OC training enhances neural processes that can be impaired in human cognitive disorders. Tasks such as OCR, VIOR, and CMOR probe how animals integrate sensory information about object features to build, maintain and utilize object representations. Supporting the translational value of using these tasks, deficits in feature binding and multisensory integration can be observed in human disorders such as AD [43], schizophrenia [44], and autism spectrum

disorder [45, 46], and have also been observed in animal models of these disorders [47-49]. In particular, the binding of object features has been shown to be impaired in the early stages of AD [50], but preserved with healthy aging in humans [51] and rodents [52]. Accordingly, neuropathological, structural and electrophysiological studies indicate impaired inter-regional communication in AD; for example, neuropathology is concentrated in areas of cortico-cortical and neocortical-hippocampal exchange as well as heteromodal association cortices [28]. Altogether, OC enrichment in the current study appeared to enhance neural processes in rodents that are at-risk in aging or AD, as well as other human disorders of cognition characterized by dysfunction in distributed brain networks.

Finally, a few clarifications regarding the present study are worth noting. First, the Standard Housing group performed significantly better compared to the Control Course group on the CMOR task. This could possibly be explained by potential adverse effects of removal of physical exercise from the CC group which was not experienced by the SH group. Rather, the SH group only experienced positive changes as it was moved into slightly larger cages post completion of the enrichment protocols. Secondly, while group means trended towards improved performance by the OC group compared to all others on several aspects of PD task acquisition, we did not find robust differences in performance between OC and EH. This is consistent with a prior report of EE primarily enhancing reversal learning compared to aspects of the acquisition phase in this task [53]. This could be related to the longitudinal nature of testing required in the present study; all groups may have benefitted as testing continued following the explicit enrichment phase [54]. Alternatively, it is possible that OC training produces differential effects on different brain circuitry, the requirements of which vary across behavioral tests—this is an intriguing possibility that requires further investigation. Indeed, additional work is needed to clarify the specific neurobiological effects of OC and CC training and how these might differ from traditional enriched housing conditions. Lastly, due to the labor intensive nature of running each animal daily for an hour, we were limited in the number of mice

per group (n = 10 mice, n = 2 groups, 20-hours per day for 20 mice). Given this practical limitation, and the abundance of male-only studies in the EE literature, we opted to test male mice only in this initial study. While we predict that female mice would undergo similar OC-mediated cognitive enhancements, future work is needed to test this important question directly.

In addition to new accounts for enrichment effects, we also propose a novel paradigm that is methodologically advantageous to existing enrichment paradigms. With current enrichment methods, it is not possible to have experimental control over the level of animal engagement with enrichment stimuli. This poses an issue when delivering enrichment with other experimental manipulations (genetic line, drug intervention, sex), which may modulate animal engagement level with enrichment stimuli. Conducting such manipulations, however, is crucial for maximizing the translational value of EE studies. With the OC, enrichment level can be matched across animals by fixing the duration spent on the course (as done in the current study) or the number of laps completed. Additionally, contrary to prior attempts of standardization [30, 31], the set-up of the course requires that the rodent interacts with each obstacle before proceeding through the track. Lastly, the OC has a control counterpart which approximates the degree of physical demand, since the CC is identical in size, contains elevated hurdles recruiting muscular activity and energy expenditure, and both the control and experimental groups were run for a fixed duration. This is contrary to running wheels that do not systematically match the amount of exercise to that achieved in the enrichment cages. Future work could further advance the design by fully automating the course via motion-sensing to enable unidirectional movement and automated reward delivery.

Altogether, the present study provides a stepping-stone to better investigations of the therapeutic potential of enriched experience. Advancing knowledge of the effects of enrichment can provide a basis for lifestyle changes and improved therapies for disease states. Moreover, separating the biological effects of different components of EE is of importance because certain patient populations experiencing cognitive decline may also have physical inabilities (e.g.,

paralysis, motor impairment) and cannot benefit from the protective effects of physical activity-based EE, but could achieve alternative forms of cognitive stimulation. Therefore, there is clinical value to understanding the cognitive enhancements of EE, and more generally, expanding upon the current methods of studying enrichment effects in rodent models.

2.5 Methods

Overview

The study was composed of four groups of mice (n = 10 per group): obstacle course (OC), control course (CC), enrichment housing (EH), and standard housing (SH) groups. The course groups, both OC and CC, were housed the same as the SH group, but were removed to run on their corresponding courses for 1 h/day, 6 days/week, for 9 weeks. Meanwhile, the EH mice remained in their enrichment cages, which were matched for object number with those of the OC. After the 9-week enrichment/control period, all animals were housed in cages intermediate in size to the SH and EH conditions without added enrichment while cognitive testing continued.

Animals

A total of 40 28-day-old male C57BL/6 mice were obtained from Charles River (St. Constant, Quebec, Canada). All animals were 42 days old at the initiation of the study. After implementation of the 9-week enrichment protocols with their corresponding controls, animals were tested on a battery of object recognition and learning tasks beginning at 4 months of age. Tests were executed monthly from 4-10 months (Table 1).

Throughout the length of experimental testing, animals in the OC, CC and SH group were housed in clear polyethylene cages (16cm x 12cm x 26 cm) in groups of 3-4. Animals in the EH group were housed in larger cages (60cm x 60cm x 25 cm) in groups of 5. All cages

were furnished with corncob bedding, crink-l'Nest and cotton nest squares for maintenance of the quality of health of the mice. Animals were fed a lightly restricted diet of 3g of food per a day (Teklad Global 18% Protein Rodent Maintenance Diet, Harlan Teklad, WI) until their weight stabilized (6 weeks on the tracks).

Subsequently, all groups were transitioned to a limited diet of 2.5g of the same food per day per mouse. Water was available for all groups ad libitum. Due to aggression, 8 out of 10 EH mice had to be individually housed immediately after the enrichment period, while only 1 out of 10 animals from the OC group was individually housed 5 months post discontinuation of enrichment. Animals were maintained on a 12-hour light-dark cycle, and tested during the light phase. All procedures followed the guidelines of the Canadian Council on Animal Care and were approved by the University of Guelph Animal Care Committee.

Course set-up

The base tracks were custom-made at the University of Lethbridge and were identical for both the OC and CC groups (Fig. 1a, c). The track is a square-shape design with the following dimensions: 860 mm arm length (including the corners) and 610 mm (excluding the corners), 120 mm arm width with surrounding walls of 70mm in height on each side. The walls were utilized to contain the animal within the course. The track was mounted on four legs 18-cm in length. The track exterior was entirely enclosed by corrugated plastic cardboard, blocking the animal's view of the testing room and limiting it to the interior of the course only. Corrugated barriers were also inserted at each corner to prevent the animal from cutting corners, ensuring passage through the course interior. Beam sensors were inserted in each corner and connected to a programmed Arduino board to track the time spent in each arm and the total time to complete each lap. Lastly, video monitoring was implemented using a top-down view of the mice for video recording. Offline video analysis using DeepLabCut was implemented for animal tracking (Supplemental Videos 3-4). Grey rectangular board pieces made of sintra foam sheets

were used as a base for obstacles, and a total of 12 inserts (base+obstacle) filled the course (3 inserts per arm) at a given time. A 0.03ml drop of Neilson ® strawberry milkshake (fat/lipid content: 22.58 g/L, sugar content: 125.81 g/L) was given as reward at the end of each lap for both OC and CC groups. A wall divider was placed immediately before the start site of the course to ensure unidirectional animal movement.

Obstacle Course protocol

Each arm of the track was allocated 3 obstacles positioned serially. Handmade obstacles were glued to the grey plastic bases to fit within the course perimeter. Daily, within each group (OC vs. CC), mice were run in a randomized order (between first to last), for a 1-hour duration over the course of 9 weeks. A detailed protocol was followed for each week, which ensured matched course alterations across animals. The protocol can be found in the supplementary material (Supplementary Figure 1) and is briefly outlined here.

All groups were habituated to handling and the OC and CC groups to food reward (strawberry milkshake) for one week. In the following week, animals were gradually habituated to an empty course for the first four days. In the subsequent 2 days, 12 triangular hurdles were gradually inserted to fill the entire course (Fig. 1c, Supplementary Figure 1). Course enrichment began after empty course and hurdle habituation. In the first week of enrichment, 3 obstacles were added each day for 4 days until the course contained 12 obstacles. In the subsequent days for 8 weeks, the following daily manipulations were implemented: change to existing obstacle→ add new obstacle → change to existing obstacle → add new obstacle → change to existing obstacle. This led to a daily protocol of adding 2 new obstacles and making 3 rearrangements to the track (Fig. 1e, Supplementary Figure 1). Changes consisted of rotating obstacles 180 degrees in their position or swapping the positions of a given pair of obstacles. These manipulations were spread across the 1-hr session, at intervals of 10 min. Supplemental Video 2 illustrates a segment of an OC session with obstacle manipulations. OC mice experienced a

total of 132 obstacles in the 9-week enrichment period (Supplemental Video 1), and 12 identical hurdles in the week prior to enrichment.

Control Course protocol

The CC group was habituated in the same manner as described above. Like the OC, the CC was gradually filled with 12 hurdles. There were no subsequent manipulations performed to the course. The CC group ran the 12-hurdle filled course for a matched duration of 1-hour for the remainder of the study. This group served as a control for physical exercise.

Home cage enrichment protocol

Larger sized cages (60cm x 60cm x 25 cm) were used for the EH group. Home-cage enrichment began on the same day as course enrichment. The addition of new objects followed the same daily pattern as the addition of new obstacles for the OC group. Three objects were added to cages daily for 3 days, followed by replacement of 2 old objects with 2 new ones. The EH group experienced a total of 132 objects throughout the duration of the enrichment period. Kraft Smooth peanut butter was put in different locations in the enriched cages to facilitate foraging and interaction with enrichment objects. In contrast, the SH cage enrichment was limited to nesting material with no additional objects or cage manipulations.

Behavioral testing

Mice were tested on six behavioral protocols, five of which exploit rodents' tendency to explore novel items preferentially [53, 55, 56] and the sixth involved a touch-screen based learning and reversal task [55]. All tests except the object location and touchscreen tasks, were conducted using an enclosed Y-shaped apparatus [37]. Each object was used once for each object-based task. Table 1 indicates the timeline for each behavioral test. Objects were always washed with ethanol immediately prior to being placed in the apparatus. These tasks were

chosen to assess different, complementary aspects of object and spatial cognitive processing and learning, partly as an index of the generalizability of the beneficial effects of the enrichment protocols. The principal behavioural score for object-based tests was an object discrimination ratio (DR), the difference in time spent exploring the novel versus the familiar object divided by the sum (DR = time of novel object exploration – time of familiar object exploration / total time of exploration). DR values were calculated for each mouse and then averaged across mice.

Spontaneous object recognition (SOR)

The SOR task [47] was used to assess whether EE mediated memory enhancements for object identity can be induced with the OC experience. SOR testing was initially implemented during the 9th week of enrichment and again 1-month following discontinuation of enrichment. The task consisted of a sample (7 min or 10 min) phase, during which animals were presented with two identical objects in each arm of the Y-apparatus (Fig. 2a). This was followed by a delay, and then the choice phase (3 min), in which one of the two objects from the sample phase was replaced with a novel object. Two delay durations were tested for each mouse, 20-min and 24-h, in order to assess short and long-term memory, respectively. The novel object position was counterbalanced between the right and left arms of the Y-apparatus.

Object-location memory (OLM)

The OLM task was used to test whether OC enrichment was sufficient to induce enhancements of spatial memory, independently of object identity [47] OLM testing was implemented 5 and 6 months post enrichment, and consisted of sample (10 or 7 min), delay (20 min or 24 h), and choice phases (3 min) (Fig. 3a). The task is run in an open field (45 × 45 x 30cm). In the sample phase, two identical objects are presented in two corners of the open field. Following the retention delay, the same identical objects are presented, but one is displaced to

the opposite corner. Preference for this displaced object indicates recognition of the original spatial location.

2-Choice pairwise visual discrimination (PD) with reversal

The PD task was used to replicate EE enhancements in cognitive flexibility [36] and to assess learning capacity [55] using a non-object-based task. This is a 10-stage protocol, with the first 7 stages aimed at gradually habituating the mouse to the experimental set-up (Supplementary Figure 2), and stages 8-10 involved learning image-reward associations as well as the reversal of such associations (Fig. 4a). Each mouse was required to reach a pre-defined criterion to proceed to the subsequent stage in the experiment. Stage 1 is a 10 min habituation to the chamber. Stage 2 is a 20 min habituation including habituation to simultaneous light, tone, and strawberry milkshake reward delivery. Stage 3 is the same as 2, with an extension to 40 min. Stage 4 involved delivering reward/tone/light upon mouse touching the singly presented stimulus. Stage 5 required the mouse to touch the presented image for reward/tone/light delivery. If criterion is not reached after a maximum of 7 sessions, the animal is taken back to stage 4. Stage 6 required the mouse to initiate a subsequent trial via a nose poke of the food tray. For stages 4-6, criterion was set as completion of 30 trials in a 60-min duration, with a total of 40 images utilized. Stage 7 introduced a punishment for touching the incorrect location, where a 5-sec time out was introduced with the chamber light turned on before the ITI and subsequent trial could be initiated. Criterion was set as completion of 24 out of 30 trials or better within a 60-min duration for 2 consecutive sessions, and two stimuli were utilized. Stages 8-9 involved presentation of both stimuli, delivery of reward upon animal contacting an S+ stimulus, and punishment following the S- stimulus contact (Fig. 4a.i). Correct behavior is followed by a new trial, and incorrect behavior is followed by repetition of the same trial until the correct response is achieved. Criterion was set as completion of 24 out of 30 trials or better within 60 min for 2 consecutive sessions. Stage 10 involved contingency reversal where responding to

the previously learned S- becomes associated with reward (S+) and vice versa (Fig. 4a.ii). The session ended after completion of 30 trials or a 60-min duration.

Object category recognition (OCR)

The OCR task was used to evaluate the ability to generalize within object categories and distinguish between separate categories [57]. OCR testing was implemented between 1 and 2-months post-discontinuation of enrichment, and consisted of sample (10 min), delay (30 min, 1 h), and choice phases (3 min) (Fig. 5a). In the sample phase, mice were exposed to two different objects of the same category (e.g., toy cars). In the choice phase, two new objects were introduced: a new object from the same category, and a new object from a different category. Preferential exploration of the new object from the previously unsampled category was interpreted as within-category generalization (i.e., a preference for the novel category).

View invariant object recognition (VIOR)

The VIOR task was used to assess flexible use of object representations when mice were presented with previously explored objects from novel perspectives [47]. VIOR testing was implemented 3-months post-discontinuation of enrichment, and consisted of sample (10 min), delay (5 min), and choice phases (3 min) (Fig. 5c). In the sample phase, two identical objects were shown in the same orientation. For the choice phase, one object was rotated (90 degrees) while the other remained in its original orientation. In the case of the VIOR task, absence of preference for either object in the choice phase is interpreted as recognition that the two objects are the same despite their differing orientations.

Cross-modal object recognition (CMOR)

The CMOR task was utilized to test object recognition across sensory modalities [47]. CMOR testing was implemented 4-months post-discontinuation of enrichment, and consisted of

sample (10 min), delay (30-min, 1 h), and choice phases (2 min) (Fig. 6a). In the tactile sample phase, mice were exposed to two identical objects, one in each arm, under red light (to prevent visual processing). In the visual choice phase, transparent barriers were inserted in each arm between the mouse and the object, preventing tactile interaction. Preference for the novel object in the choice phase thus indicates cross-modal (tactile-to-visual) object recognition.

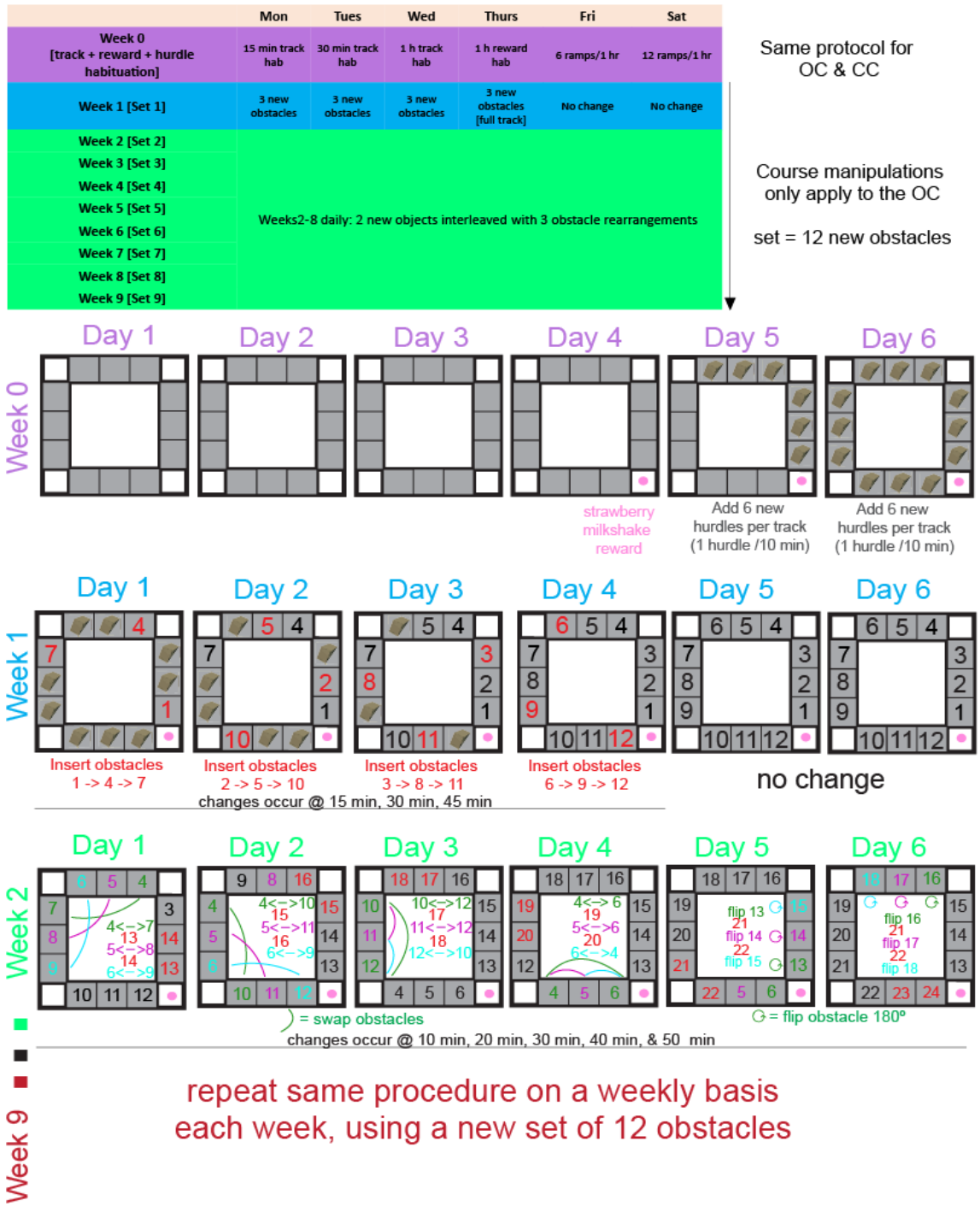
General behavioral procedure and statistical analysis

Regarding the OR tasks, the behavioral testing room, Y-maze apparatus, video-scoring procedure and statistical analyses were identical to those described in Creighton et al., 2019. To clarify further, for all object tasks, a DR was calculated for both the sample and choice phases. The sample phase DR reflects exploration when the objects are equally novel. Using paired samples t-tests, the choice phase DR was then compared to the sample phase DR to assess whether mice discriminated significantly above chance. Additionally, for these tasks, each animal underwent an individual trial. However, if high variance was observed, each animal experienced a second trial, where a single mean value across the two trials was obtained per animal.

The visual discrimination and reversal tasks were conducted in automated touchscreen operant chambers (Bussey-Saksida Mouse Touchscreen Chamber System, Lafayette Instrument Co., Lafayette, IN), with instructions and event recordings operated through the software Whisker Server and ABET II.

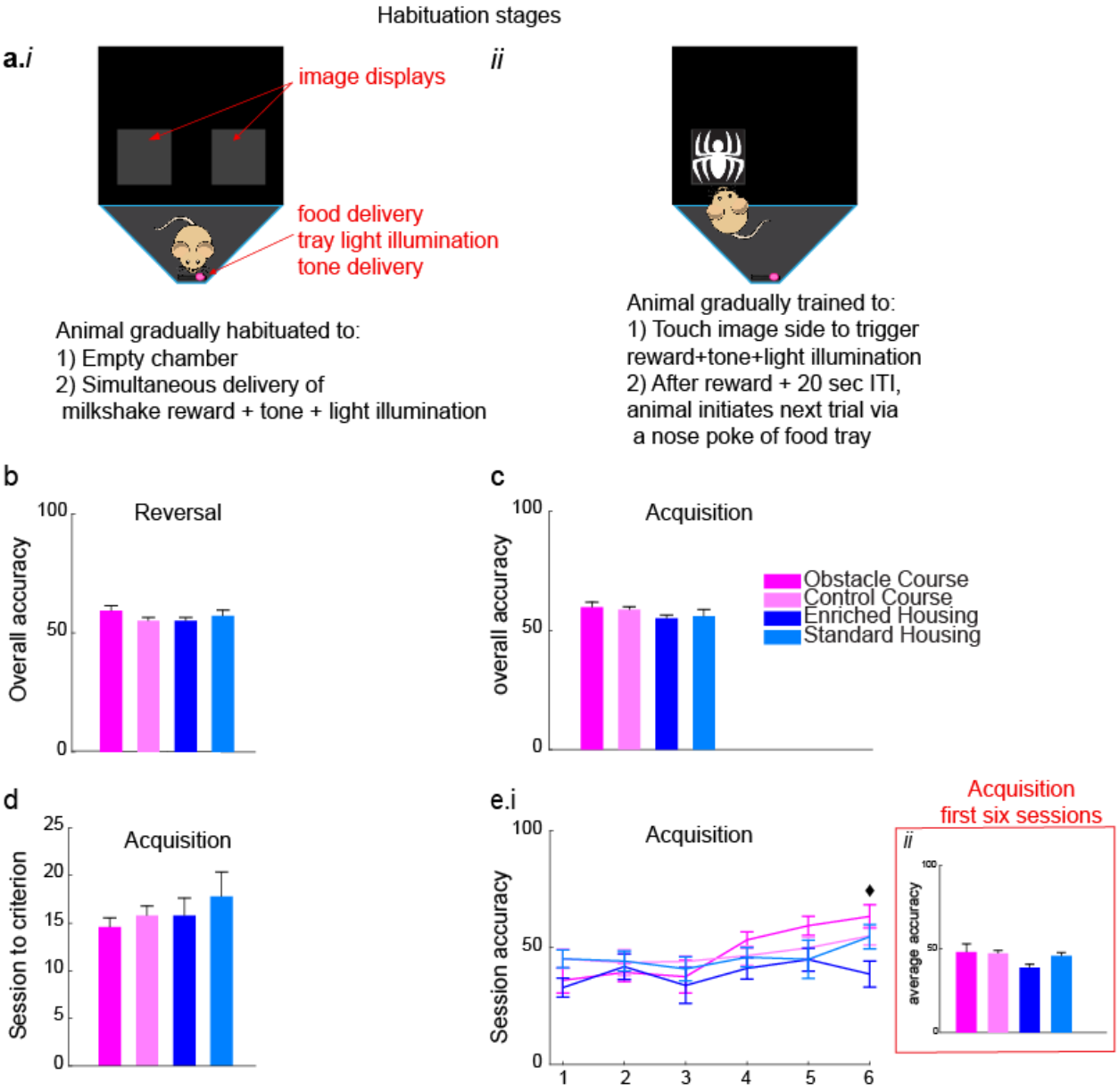
For PD task statistical analysis, a repeated-measures four-way analysis of variance (ANOVA) was used to analyze between-group differences across the relevant testing days. Based on significant differences identified, appropriate *post-hoc* analysis was performed sequentially. All other group comparisons were completed with a one-way ANOVA and appropriate *post-hoc* tests.

2.6 Supplementary material



Supplementary Figure 1. Detailed OC enrichment protocol
Top panel is a table outlining OC and CC protocols including weekly and daily executions. Bottom schematics further detail the daily and weekly manipulations to the track. Habituation occurred at week 0 in the same manner for both the OC and CC groups. Weeks 1 – 9 apply to the OC group only. By the end of week 1, the OC was filled with novel obstacles. The manipulations that took place on week 2 were done in a similar manner for weeks 3 - 9. By the

end of each week, a full new set of 12 obstacles were introduced to the OC track. Enrichment occurred for a duration of 9 weeks. Each day, mice within each group were run on their respective tracks in a randomized order between one and 10 (first to last).



Supplementary Figure 2. PD task – habituation schematic and supplementary results
a. Schematic representation of the PD habituation stages. **a.i-ii.** Gradually habituating the animal to: **i**. recording chamber, simultaneous delivery of reward, tone, and light, **ii.** touching a presented stimulus to trigger reward, and nose-poke in reward tray to trigger stimulus presentation. **b-c.** Mean accuracy across all **(b)** reversal $F(3, 36) = 1.196$, $p = 0.325$, partial $\eta^2 = 0.091$ and **(c)** acquisition $F(3, 36) = 1.945$, $p = 0.14$, partial $\eta^2 = 0.139$ sessions per group. **d.** Mean session number to reach criterion (performance of 80% accuracy on two successive sessions) per group in the acquisition phase. No group main effect was observed $F(3,36) =$

2.004, p = 0.131, partial η^2 = 0.143. **e.i.** Group percent accuracy on the first six acquisition sessions with a main effect of session, F(5, 36) = 6.975, p < 0.001, partial η^2 = 0.162 and an effect of group in session six, F(3,36) = 4.340, p = 0.010, partial η^2 = 0.266, with the OC group performing significantly higher than EH on session six t(18) = 3.294, p = 0.004. ***ii.*** Mean accuracy across the first six acquisition sessions for each group revealed no significant differences in group accuracy, F(3,36) = 1.78, p = 0.168, partial η^2 = 0.129.

Supplementary Video 1. Example obstacles
A video displaying utilized obstacles for the OC.

Supplementary Video 2. Example obstacle course session
Example Obstacle Course session from week 5, day 1, animal # 13. Demonstrated in the video are 3 consecutive runs on the OC by the same mouse: one lap prior to and two laps following course manipulation. As demonstrated, animals tended to take longer time to complete laps immediately following course manipulation, spending more time interacting with the manipulated or new obstacle. However, animal speed returned to baseline on the second lap after manipulation. In this particular video, the manipulation involved swapping two already existing obstacles.

Supplementary Video 3-4. Deep learning tracking of animal performance on the OC (Video 3) and CC (Video 4)
Using image frames extracted from video monitoring of animals on the course, DeepLabCut was trained to track animal position (cyan circle) in order to calculate time to complete laps on the respective tracks.

2.7 References

1. DeFelipe, J., *Brain plasticity and mental processes: Cajal again.* Nat Rev Neurosci, 2006. **7**(10): p. 811-7.
2. Mora, F., G. Segovia, and A. del Arco, *Aging, plasticity and environmental enrichment: structural changes and neurotransmitter dynamics in several areas of the brain.* Brain Res Rev, 2007. **55**(1): p. 78-88.
3. Morris, R.G., *D.O. Hebb: The Organization of Behavior, Wiley: New York; 1949.* Brain Res Bull, 1999. **50**(5-6): p. 437.
4. Rosenzweig, M.R., et al., *Effects of environmental complexity and training on brain chemistry and anatomy: a replication and extension.* J Comp Physiol Psychol, 1962. **55**: p. 429-37.
5. Rosenzweig, M.R. and E.L. Bennett, *Psychobiology of plasticity: effects of training and experience on brain and behavior.* Behav Brain Res, 1996. **78**(1): p. 57-65.
6. van Praag, H., G. Kempermann, and F.H. Gage, *Neural consequences of environmental enrichment.* Nat Rev Neurosci, 2000. **1**(3): p. 191-8.
7. Jankowsky, J.L., et al., *Environmental enrichment mitigates cognitive deficits in a mouse model of Alzheimer's disease.* J Neurosci, 2005. **25**(21): p. 5217-24.
8. Laviola, G., et al., *Effects of enriched environment on animal models of neurodegenerative diseases and psychiatric disorders.* Neurobiol Dis, 2008. **31**(2): p. 159-68.
9. Segovia, G., A. del Arco, and F. Mora, *Environmental enrichment, prefrontal cortex, stress, and aging of the brain.* J Neural Transm (Vienna), 2009. **116**(8): p. 1007-16.

10. Pang, T.Y. and A.J. Hannan, *Enhancement of cognitive function in models of brain disease through environmental enrichment and physical activity.* Neuropharmacology, 2013. **64**: p. 515-28.
11. Paillard, T., Y. Rolland, and P. de Souto Barreto, *Protective Effects of Physical Exercise in Alzheimer's Disease and Parkinson's Disease: A Narrative Review.* J Clin Neurol, 2015. **11**(3): p. 212-9.
12. Birch, A.M. and A.M. Kelly, *Lifelong environmental enrichment in the absence of exercise protects the brain from age-related cognitive decline.* Neuropharmacology, 2019. **145**(Pt A): p. 59-74.
13. Milgram, N.W., et al., *Neuroprotective effects of cognitive enrichment.* Ageing Res Rev, 2006. **5**(3): p. 354-69.
14. Lazarov, O., et al., *Environmental enrichment reduces Abeta levels and amyloid deposition in transgenic mice.* Cell, 2005. **120**(5): p. 701-13.
15. Bosma, H., et al., *Mental work demands protect against cognitive impairment: MAAS prospective cohort study.* Exp Aging Res, 2003. **29**(1): p. 33-45.
16. Kramer, A.F., et al., *Environmental influences on cognitive and brain plasticity during aging.* J Gerontol A Biol Sci Med Sci, 2004. **59**(9): p. M940-57.
17. Clemenson, G.D., C.M. Henningfield, and C.E.L. Stark, *Improving Hippocampal Memory Through the Experience of a Rich Minecraft Environment.* Front Behav Neurosci, 2019. **13**: p. 57.
18. Cheng, S.T., *Cognitive Reserve and the Prevention of Dementia: the Role of Physical and Cognitive Activities.* Curr Psychiatry Rep, 2016. **18**(9): p. 85.
19. Simpson, J. and J.P. Kelly, *The impact of environmental enrichment in laboratory rats--behavioural and neurochemical aspects.* Behav Brain Res, 2011. **222**(1): p. 246-64.
20. Bennett, J.C., et al., *Long-term continuous, but not daily, environmental enrichment reduces spatial memory decline in aged male mice.* Neurobiol Learn Mem, 2006. **85**(2): p. 139-52.
21. Bernstein, L., *A study of some enriching variables in a free-environment for rats.* J Psychosom Res, 1973. **17**(2): p. 85-8.
22. Frick, K.M. and S.M. Fernandez, *Enrichment enhances spatial memory and increases synaptophysin levels in aged female mice.* Neurobiol Aging, 2003. **24**(4): p. 615-26.
23. Garthe, A., I. Roeder, and G. Kempermann, *Mice in an enriched environment learn more flexibly because of adult hippocampal neurogenesis.* Hippocampus, 2016. **26**(2): p. 261-71.
24. Rampon, C., et al., *Enrichment induces structural changes and recovery from nonspatial memory deficits in CA1 NMDAR1-knockout mice.* Nat Neurosci, 2000. **3**(3): p. 238-44.
25. Gortz, N., et al., *Effects of environmental enrichment on exploration, anxiety, and memory in female TgCRND8 Alzheimer mice.* Behav Brain Res, 2008. **191**(1): p. 43-8.
26. Jeyaraj, S.E., et al., *Environmental enrichment enhances sociability by regulating glutamate signaling pathway through GR by epigenetic mechanisms in amygdala of Indian field mice Mus booduga.* Gen Comp Endocrinol, 2021. **300**: p. 113641.
27. Roberts, K.L. and H.A. Allen, *Perception and Cognition in the Ageing Brain: A Brief Review of the Short- and Long-Term Links between Perceptual and Cognitive Decline.* Front Aging Neurosci, 2016. **8**: p. 39.
28. Delbeuck, X., M. Van der Linden, and F. Collette, *Alzheimer's disease as a disconnection syndrome?* Neuropsychol Rev, 2003. **13**(2): p. 79-92.
29. Murman, D.L., *The Impact of Age on Cognition.* Semin Hear, 2015. **36**(3): p. 111-21.
30. Fares, R.P., et al., *Standardized environmental enrichment supports enhanced brain plasticity in healthy rats and prevents cognitive impairment in epileptic rats.* PLoS One, 2013. **8**(1): p. e53888.

31. Fares, R.P., H. Kouchi, and L. Bezin, *Standardized environmental enrichment for rodents in Marlau cage.* Protocol Exchange, 2012.
32. Birch, A.M., N.B. McGarry, and A.M. Kelly, *Short-term environmental enrichment, in the absence of exercise, improves memory, and increases NGF concentration, early neuronal survival, and synaptogenesis in the dentate gyrus in a time-dependent manner.* Hippocampus, 2013. **23**(6): p. 437-50.
33. Mesa-Gresa, P., A. Perez-Martinez, and R. Redolat, *Environmental enrichment improves novel object recognition and enhances agonistic behavior in male mice.* Aggress Behav, 2013. **39**(4): p. 269-79.
34. Frick, K.M., et al., *Effects of environmental enrichment on spatial memory and neurochemistry in middle-aged mice.* Learn Mem, 2003. **10**(3): p. 187-98.
35. Roebuck, A.J., et al., *Cognitive Impairments in Touchscreen-based Visual Discrimination and Reversal Learning in Genetic Absence Epilepsy Rats from Strasbourg.* Neuroscience, 2020. **430**: p. 105-112.
36. Zeleznikow-Johnston, A., et al., *Environmental enrichment enhances cognitive flexibility in C57BL/6 mice on a touchscreen reversal learning task.* Neuropharmacology, 2017. **117**: p. 219-226.
37. Creighton, S.D., et al., *Development of an "object category recognition" task for mice: Involvement of muscarinic acetylcholine receptors.* Behav Neurosci, 2019. **133**(5): p. 527-536.
38. McClelland, J.L., *Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory.* J Exp Psychol Gen, 2013. **142**(4): p. 1190-1210.
39. McClelland, J.L., B.L. McNaughton, and A.K. Lampinen, *Integration of new information in memory: new insights from a complementary learning systems perspective.* Philos Trans R Soc Lond B Biol Sci, 2020. **375**(1799): p. 20190637.
40. Tse, D., et al., *Schemas and memory consolidation.* Science, 2007. **316**(5821): p. 76-82.
41. Stringer, S.M. and E.T. Rolls, *Invariant object recognition in the visual system with novel views of 3D objects.* Neural Comput, 2002. **14**(11): p. 2585-96.
42. Zoccolan, D., *Invariant visual object recognition and shape processing in rats.* Behav Brain Res, 2015. **285**: p. 10-33.
43. Martinez, J.F., et al., *Assessment of Conjunctive Binding in Aging: A Promising Approach for Alzheimer's Disease Detection.* J Alzheimers Dis, 2019. **69**(1): p. 71-81.
44. Williams, L.E., et al., *Reduced multisensory integration in patients with schizophrenia on a target detection task.* Neuropsychologia, 2010. **48**(10): p. 3128-36.
45. Baum, S.H., R.A. Stevenson, and M.T. Wallace, *Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder.* Prog Neurobiol, 2015. **134**: p. 140-60.
46. Stevenson, R.A., et al., *Conjunctive Visual Processing Appears Abnormal in Autism.* Front Psychol, 2018. **9**: p. 2668.
47. Creighton, S.D., et al., *Dissociable cognitive impairments in two strains of transgenic Alzheimer's disease mice revealed by a battery of object-based tests.* Sci Rep, 2019. **9**(1): p. 57.
48. Gogolla, N., et al., *Sensory integration in mouse insular cortex reflects GABA circuit maturation.* Neuron, 2014. **83**(4): p. 894-905.
49. Jacklin, D.L., et al., *Severe cross-modal object recognition deficits in rats treated sub-chronically with NMDA receptor antagonists are reversed by systemic nicotine: implications for abnormal multisensory integration in schizophrenia.* Neuropsychopharmacology, 2012. **37**(10): p. 2322-31.

50. Parra, M.A., et al., *Brain Information Sharing During Visual Short-Term Memory Binding Yields a Memory Biomarker for Familial Alzheimer's Disease.* Curr Alzheimer Res, 2017. **14**(12): p. 1335-1347.
51. Rhodes, S., M.A. Parra, and R.H. Logie, *Ageing and feature binding in visual working memory: The role of presentation time.* Q J Exp Psychol (Hove), 2016. **69**(4): p. 654-68.
52. Gaynor, L.S., et al., *Impaired discrimination with intact crossmodal association in aged rats: A dissociation of perirhinal cortical-dependent behaviors.* Behav Neurosci, 2018. **132**(3): p. 138-151.
53. Winters, B.D., et al., *Double dissociation between the effects of peri-postrhinal cortex and hippocampal lesions on tests of object recognition and spatial memory: heterogeneity of function within the temporal lobe.* J Neurosci, 2004. **24**(26): p. 5901-8.
54. Mallien, A.S., et al., *Daily exposure to a touchscreen-paradigm and associated food restriction evokes an increase in adrenocortical and neural activity in mice.* Horm Behav, 2016. **81**: p. 97-105.
55. Horner, A.E., et al., *The touchscreen operant platform for testing learning and memory in rats and mice.* Nat Protoc, 2013. **8**(10): p. 1961-84.
56. Winters, B.D., L.M. Saksida, and T.J. Bussey, *Object recognition memory: neurobiological mechanisms of encoding, consolidation and retrieval.* Neurosci Biobehav Rev, 2008. **32**(5): p. 1055-70.
57. Winters, B.D. and J.M. Reid, *A distributed cortical representation underlies crossmodal object recognition in rats.* J Neurosci, 2010. **30**(18): p. 6253-61.

Chapter 3: Rodent behavior and associated neural activity revealed through spectral analysis and machine learning (deductive):

A role for proximal CA1 20-40 Hz dynamics for performance on a nonspatial sequence memory task

1.1 Abstract

The hippocampus is known to play a critical role in processing information about temporal context. However, it remains unclear how hippocampal oscillations are involved, and how their functional organization is influenced by connectivity gradients. We examined local field potential activity in CA1 as rats performed an odor sequence memory task. We found that odor sequence processing epochs were characterized by distinct spectral profiles and proximo-distal CA1 gradients of theta and 20-40 Hz power compared to running epochs. We found that 20-40 Hz power was predictive of sequence memory performance, particularly in proximal CA1 and during the plateau of high power observed on trials in which animals had to maintain their decision until instructed to respond. Altogether, these results provide evidence that dynamics of 20-40 Hz power along the CA1 axis support nonspatial information processing critical for performance on an odor-based sequence memory task and are consistent with a role for 20-40 Hz power in gating information processing.

3.2 Introduction

Frequency content in local field potential recordings is preserved across species, varies predictably with behavior, and is central to brain function (Buzsaki and Watson, 2012). Across a broad range of frequencies (0.5 – 500 Hz), spectral features including power, phase, coupling thereof, and oscillatory modulation of spiking activity have been shown to be biologically meaningful, and are thought to be a means of effective inter and intra regional communication

(Fries 2005, Lisman, 2005, Chalk et al., 2016, Schyns et al., 2011, Buzsaki and Watson, 2012). These spectral dynamics have been observed with distinct cognitive states and can be used to decode behavior (Trimper et al., 2017, Rangel et al., 2017, Bae and Luck, 2019, Kikumoto and Mayr, 2018). Therefore, spectral dynamics hold informational value and may elucidate neural computations including those supporting learning and memory (Axmacher et al., 2006).

It is well established that the hippocampus is critical for spatial learning and memory (O'Keefe & Nadel, 1978) and that distinct frequency content is observed in hippocampal subregion CA1 during spatial navigation in rodents (see Colgin, 2016). CA1 theta power increases with running speed and exploration and its phase is thought to link representations of a spatial trajectory and, more generally, constituent items of a given episodic experience (Buzaki & Moser, 2013, Dreu and Zugaro, 2019). The CA1 also exhibits transient increases in slow (25-55 Hz) and fast gamma (60-100 Hz) power during running, through which the CA1 is coherent with the CA3 and entorhinal cortex, respectively (Colgin et al.,2009). Slow gamma was also shown to support associative memory encoding, predicting subsequent retrieval (Trimper et al., 2017). In addition, frequency dynamics during spatial memory tasks exhibit heterogeneity along the CA1 proximo-distal axis, possibly pointing to distinct coding properties associated with the underlying engaged circuits. Findings from prior studies support the notion that coding of fundamental aspects of an experience (space, context) increases towards proximal (pCA1) (Henrikson et al., 2010, Nakazawa et al., 2016) while coding of secondary associations made within space (odor-location) may increase towards distal (dCA1) (Igarashi et al., 2014). It remains unclear, however, the degree to which the aforementioned frequency content and associated gradients during spatial memory extend to memory for temporal order.

Consistent with its unique architecture and connectivity (McNaughton and Morris, 1987; Lisman 1999; Foster and Knierim 2012; Buzsáki and Tingley 2018), a growing literature shows the hippocampus is also critical for remembering sequences of nonspatial events (Fortin et al. 2002; Kesner et al. 2002; Allen and Fortin, 2013; Eichenbaum, 2014) and that hippocampal

neurons code for temporal relationships among such events (MacDonald et al., 2011; Allen et al. 2016; Shahbaba et al., 2019). The hippocampal population level spectral dynamics for sequence memory have mainly been largely understudied (Nakazono et al., 2015, and Hattori & Sakata, 2014). We previously showed that a 20-40 Hz rhythm in the CA1 is recruited during an odor-based sequence memory task (Allen et al., 2016). Activity in the 20-40 Hz range overlaps in frequency with slow gamma (~25-55 Hz) and beta (~15-35 Hz). An emerging hypothesis is that beta serves a "gating" function, based on findings from striatal and cortical circuits (Leventhal et al., 2012; Engel & Fries, 2010; Schmidt et al., 2019), whereby beta reflects a neural *status quo* bridging the period between the decision and the response. it is unclear whether this hypothesis extends to the role of 20-40 Hz in the hippocampal CA1, and more generally, the recruited spectral profiles in the CA1 which support memory for temporal order.

To address these issues, we examined CA1 local field potential recordings while rats performed a hippocampal-dependent odor-based sequence memory task (Fortin et al., 2002). This task offers precise time-locking to stimulus presentations and responses, as well as distinct trial types associated with distinct cognitive demands. As in our previous work (Allen et al., 2016), we report power increases in the 20-40 Hz and theta frequency ranges during the odor sequence processing periods. Here we extend these results by demonstrating that the same electrodes exhibited a distinct spectral content in a different state (running on the track), which was characterized by high theta power and a broadband increase in power for frequencies above 24 Hz. Power in the recruited frequencies varied along the proximodistal axis of CA1 in a behavioral state dependent manner. We also made several contributions to our understanding of 20-40 Hz oscillations to hippocampal function. We observed that 20-40 Hz power gradually increased during odor presentations and reached a plateau (steady-state) on trial types involving a *sustained hold response*. In these ''hold response" trial types, 20-40 Hz power increased with session performance, whereas oscillations in other frequency ranges showed a negative relationship with performance. Finally, we identified that 20-40 Hz steady-state power

in proximal CA1 is a key predictor of the performance-related effects we observed. Collectively, these results suggest that 20-40 Hz power dynamics in the proximal segment of CA1 are linked to processing the temporal context of nonspatial events and reflect a post-decision state serving to protect local, task-critical information processing until a response can be performed.

3.3 Results

Five rats were trained on an odor-based sequence memory task that critically depends on the hippocampus (Fortin et al., 2016). In this task, animals receive repeated presentations of a sequence of five odor items, and must correctly identify each item as being presented "in sequence" (InSeq; e.g., ABC…) or "out of sequence" (OutSeq; e.g., ABE…) to receive a water reward (Fig 1A-C). More specifically, each odor presentation is triggered by the rat's nose entering to odor port and animals need to behaviorally indicate whether the presented item is InSeq (by continuing to hold their nosepoke response until an auditory signal occurs at 1.2 sec) or OutSeq (by withdrawing their nose before the signal). The exact time cut-off to classify InSeq from OutSeq responses was determined separately per animal based on the intersection of distributions for InSeq and OutSeq poke durations (Fig 1G). This was done because while centered around 1.2 sec, there was some individual animal variability in the exact time cut-off for indicating in InSeq response. This variability is possibly due to animals relying on their own timing rather than on the auditory signal to withdraw. For InSeq items, hold durations longer and shorter than the identified time cut-off reflected correct and incorrect responses, respectively (InSeq+ and InSeq-; Fig 1C). The converse was true for OutSeq trials; poke durations shorter and longer than the time-cutoff reflected correct (OutSeq+) and incorrect (OutSeq-) responses, respectively.

Following a training period of 2-4 months, animals were surgically implanted with a microdrive over the dorsal CA1 region of the hippocamps and tetrodes were then slowly driven

to the pyramidal layer over a period of 2-3 weeks. LFP activity was then recorded from tetrodes spanning much of the CA1 proximo-distal axis as animals performed the task (Fig 1D-F). We analyzed LFP activity from three sessions: a session in which animals were tested on the sequence learned preoperatively (ABCDE; well-trained session), as well as two successive sessions testing a novel sequence (VWXYZ; novel1 and novel2 sessions). According to our previously established performance metric (Sequence Memory Index, SMI; Allen et al., 2014), performance in these sessions ranged from low, to moderate, to high, providing an opportunity to compare LFP activity across three different levels of performance.

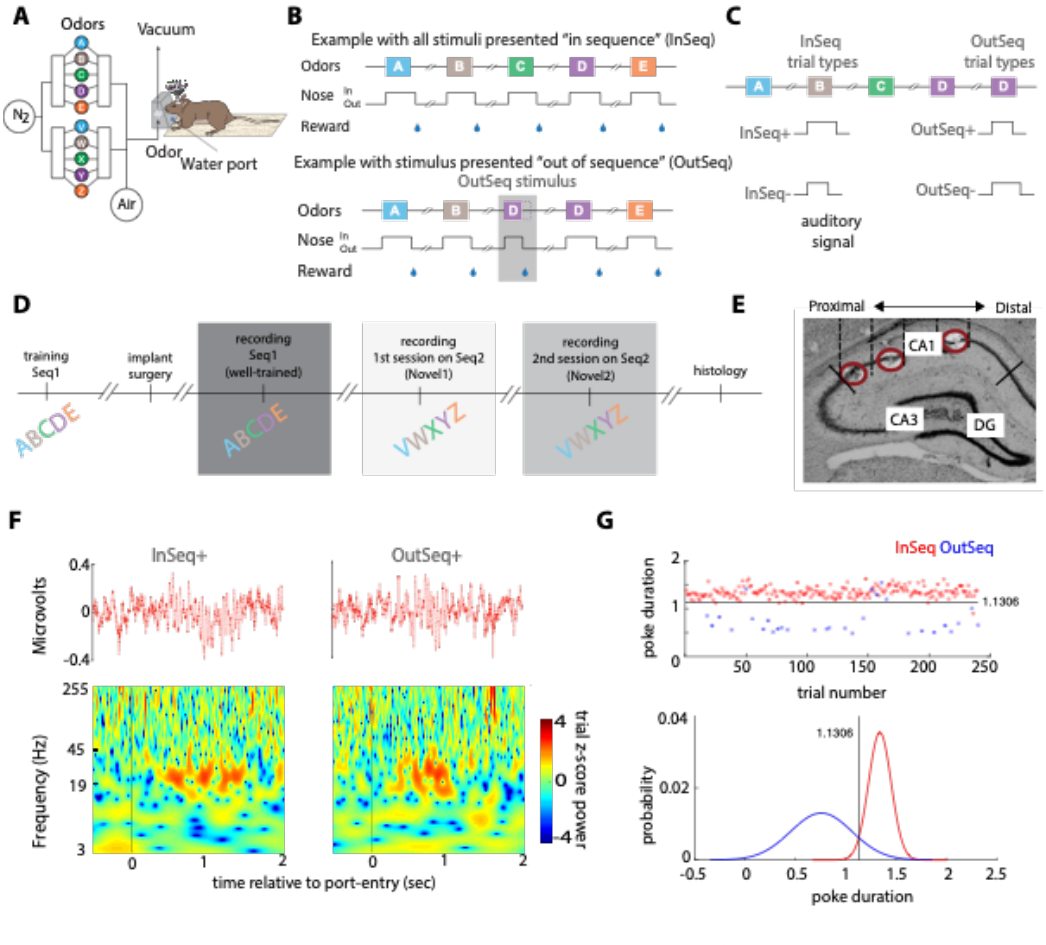


Figure 1. Experimental protocol for the investigation of CA1 oscillatory dynamics during a sequence memory task
A. Schematic of behavioral apparatus. Using an automated odor delivery system located at one end of a linear track, rats were presented with a series of five odors delivered in the same odor port. Delivery of each odor was initiated with a nose poke in. **B.** Schematic of trial presentations.

In each session, the same sequence of five odors was presented multiple times. Approximately half the presentations included all items “in sequence” (InSeq; e.g., ABCDE) (top panel) and the other half including one “out of sequence” (OutSeq; e.g., ABDDE) item (bottom panel). Odor presentation terminated with the onset of an auditory signal at 1.2 sec (indicated in B). Rats were required to behaviorally indicate whether the presented odor is InSeq (by holding their nose poke response until the auditory signal; > 1.2 s) or OutSeq (by withdrawing their nose poke before the signal; <1.2 s) to receive a water reward. After completion of each sequence (correctly or incorrectly), animals were required to run to the opposite end of the linear track and return to the odor port before the next sequence could be presented. **C.** Schematic of trial type definitions. InSeq+ (InSeq correct) and InSeq- (InSeq incorrect) trial types reflect trials where items were presented in-sequence, for which nose poke duration was longer or shorter than ~1.2 sec, respectively. OutSeq+ (OutSeq correct) and OutSeq- (OutSeq incorrect) trial types reflect trials where items were presented out-of-sequence, for which nose poke duration was shorter or longer than ~1.2 sec, respectively. **D.** Schematic of experimental timeline. Electrode hyperdrive implantation took place after animals were trained on sequence 1 (A-B-C-D-E) for 2-4 months. Following recovery, neural activity was recorded while the animal performed the already learned task (well-trained). Following data acquisition of the well-trained session, neural activity was recorded during two different stages of learning a novel odor sequence (V-W-X-Y-Z), “ novel1” and “novel2”, respectively. Histology was done after experiment termination for electrode localization. **E**. Sample histology image showing the range of tetrode tip locations, which spanned much of the proximodistal axis of CA1 (3 tip locations shown; red circles). For each animal, a set of four tetrodes equally distributed across the proximodistal axis (with comparable locations across animals) was used for local field potential activity analysis. **F.** Representative InSeq+ (left) and OutSeq+ (right) LFP data (top panel) and corresponding time-frequency plots. **G.** (top) Poke durations for all InSeq and OutSeq trials during one well-trained session for a representative animal. (bottom) Normal distribution functions fit separately for the observed InSeq and OutSeq trial poke durations in the top panel. The intersection of the two distributions (t=1.1306 sec) was used as the time cut-off to determine response accuracy (hold durations > and < 1.1306 sec reflected InSeq and OutSeq responses, respectively). Panels from this figure were adapted from our previous work (Allen, 2016).

Odor sequence processing and running epochs are associated with distinct spectral features, which vary across the CA1 proximodistal axis.

We began by investigating whether distinct spectral features in CA1 accompany odor sequence and running periods of the task, and whether the observed spectral content varied across the proximodistal axis of CA1. To do so, group (n=5) peri-event spectrograms were generated from four electrode locations along the proximodistal axis and aligned to odor processing (Fig 2A) and running (Fig 2B) epochs. For each electrode site, spectrograms were generated using analytic morlet wavelets and spectral power was z-score normalized to power at the same site during a 30-minute period of the recording session. Data were taken from the

well-trained session in which animals performed at a high level (Allen et al., 2016). For all group analyses, a random intercept linear mixed-effects model (LMEM) was fit to the data where electrodes, sessions, or trial types were treated as fixed effects and subjects were treated as a random effect (to account for inherent correlations from repeated measures at the subject level; see methods for model formulation).

We found that power in the 20-40 Hz and theta (4-12 Hz; defined *a priori*) range observed during odor presentations showed significantly distinct patterns across the proximodistal axis. More specifically, 20-40 Hz power (110-1200 msec period following port entry) increased toward proximal CA1 (LMEM ANOVA: significant effect of electrode $F_{3,2214}$ = 41.995 $p < 0.001$). Power in the most proximal site (electrode 1) showed significantly higher power compared to the two most distal sites (electrode 3 and 4) (t_{2214} = -8.072 p <0.001 and t_{2214} = -8.024 $p < 0.001$, respectively). Power in electrode 1 did not differ significantly from that of electrode 2 (t_{2214} = -0.206 p = 0.837). In contrast, theta power in the same period significantly increased towards distal CA1 (LMEM ANOVA: significant effect of electrode $F_{3,2214}$ = 60.952 $p <$ 0.001). Electrodes 2, 3 and 4 showed significantly higher theta power compared to electrode 1 (t_{2214} = 5.008 $p < 0.001$, t_{2214} = 11.99 $p < 0.001$, and t_{2214} = 10.733 $p < 0.001$, respectively).

The same electrodes exhibited a different spectral pattern during running periods, which occurred between sequence presentations (Fig 2B, E-F). More specifically, the 20-40 Hz power observed during odor sampling was weak during running. Additionally, contrary to the odor processing periods, theta power during running was significant higher in proximal CA1 (LMEM ANOVA: significant effect of electrode $F_{3,2768}$ = 71.005 $p < 0.001$) and was characterized by a higher center frequency (~ 9Hz during running compared to ~6 Hz during odor sampling). Electrode 1 showed significantly higher power compared to electrodes 3 and 4 (t_{2768} = -8.173 $p <$ 0.001, and t_{2768} = -12.414 $p < 0.001$, respectively) while power in electrode 1 did not differ significantly from that of electrode 2 (t_{2768} = -0.876 p = 0.381; Fig 2E). The running period was also characterized by increased high frequency broadband power (> 24 Hz to avoid theta's first

harmonic), which significantly increased towards distal CA1 (LMEM ANOVA: significant effect of electrode $F_{3,2768} = 181.34$, $p < 0.001$). Electrodes 3 ($t_{2768} = 13.255$ $p < 0.001$) and 4 ($t_{2768} = 19.53$ $p < 0.001$) showed significantly higher power compared to electrode 1. Power in electrode 2 did not differ significantly from that of electrode 1 ($t_{2768} = 1.051$ $p = 0.293$).

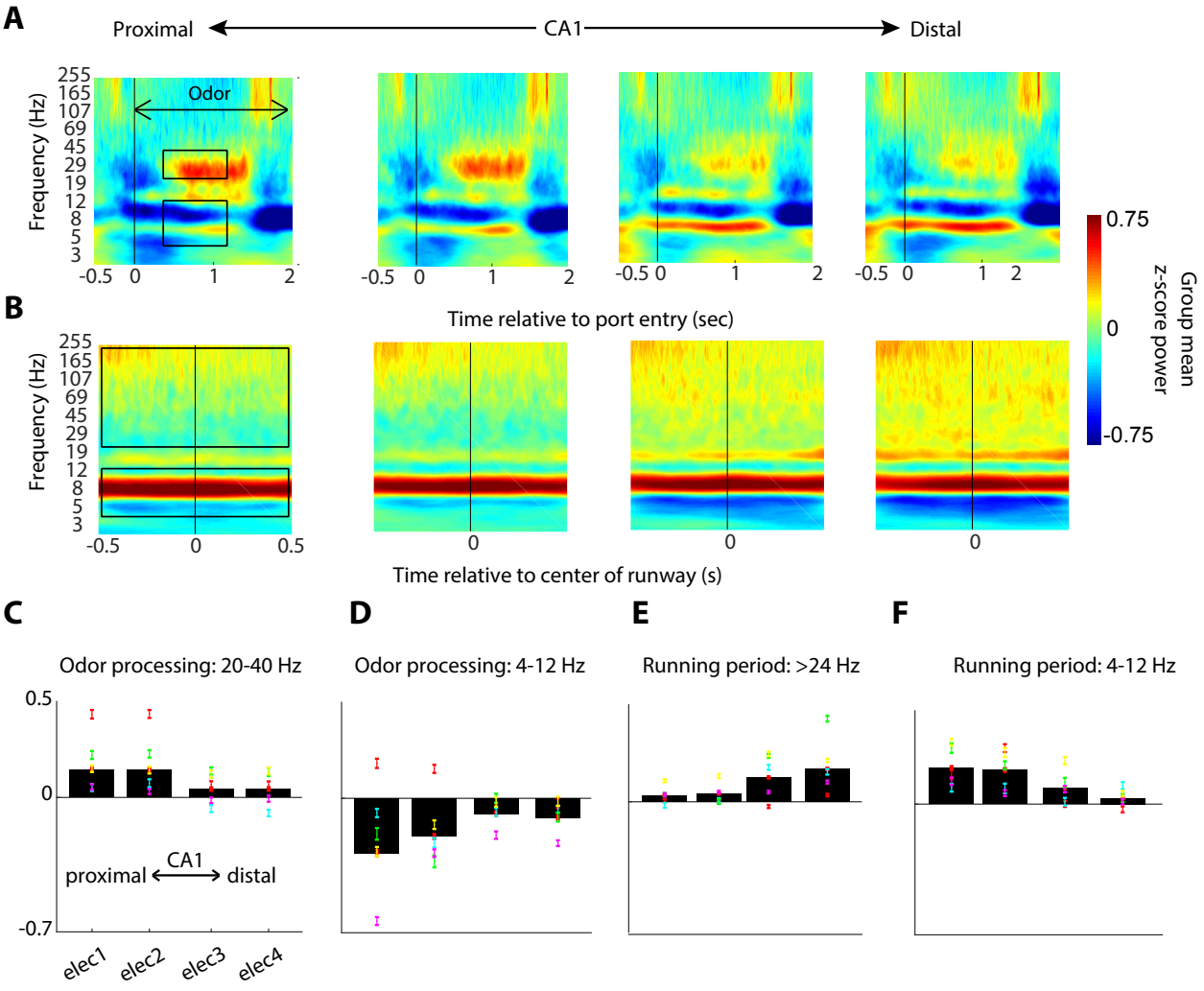


Figure 2. Odor sequence processing and running are associated with distinct CA1 oscillatory states, which vary across the proximodistal axis

A-B. Group peri-event spectrograms (n=5) during (A) odor sampling period (InSeq+ trials only) and (B) running periods in four electrode locations along the CA1 proximodistal axis (0ms = port entry). **C.** Mean z-score 20-40 Hz power during 110-1200ms period of odor presentation significantly increased toward proximal CA1 (significant effect of electrode $F_{3,2214} = 41.995$ $p = 2.1702e-26$). **D.** Same as in (C), but for theta (4-12 Hz) showing increased power towards distal

CA1 (significant effect of electrode $F_{3,2214} = 60.952$ $p < 0.001$). The time-frequency periods used in C-D are indicated by black boxs in panel A; defined *a priori*. **E.** Mean z-score higher frequency power (>24 Hz to avoid theta's first harmonic) increases along the CA1 axis (significant effect of electrode $F_{3,2768} = 181.34$, $p < 0.001$). **F.** Same as in (E), but for theta showing decreased power along the CA1 axis (significant effect of electrode $F_{3,2768} = 71.005$ $p < 0.001$). Individual animal values and within animal SEM across trials are overlaid for each barplot. All data were extracted from the well-trained recording session (high performance level). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ indicate significant differences between groups two-tailed t-test, Bonferrnoi correction $p = 0.0167$).

CA1 20-40 Hz power increases with session performance.

We then examined whether power in the 20-40 Hz range was linked to performance, and whether this association varied along the proximodistal axis. To do so, we extended the previous analyses, which were applied to the well-trained session, to two consecutive sessions in which animals learned a novel odor sequence (correctly identified InSeq trials only). This allowed us to compare power across three sessions characterized by low (novel1, first session on novel sequence), moderate (novel2, second session on novel sequence) and high levels of performance (well-trained session; Fig 3). Unlike the previous analyses, here we evaluated 20-40 Hz power during the 250 msec time window preceding the withdraw response to capture the period of high power observed toward the end of odor presentations (at which point power has plateaued; Fig 3B).

We found that 20-40 Hz power significantly increased with performance level (LMEM ANOVA: significant effect of session $F_{2,4812} = 46.71$ $p < 0.001$), which complements our previous study showing learning-related differences in waveform amplitude (Allen et al., 2016). This performance effect did not significantly vary across the proximodistal axis (non-significant interaction between electrode and session $F_{6,4806} = 0.745$ $p = 0.614$), but instead scaled with the local amplitude of 20-40 Hz, suggesting this behavioral correlate is present throughout dorsal CA1. Comparing the effect due to electrode, power is highest in electrode 1 and lowest in electrode 4 for the three sessions. Average power is significantly higher in electrode 1 compared to electrodes 2, 3, and 4 ($t_{4812} = -2.969$ $p = 0.003$, $t_{4812} = -8.896$ $p < 0.001$, and $t_{4812} =$

-11.095 $p < 0.001$, respectively), after Bonferroni correction for multiple comparisons (*corrected p-value* = 0.010) (Fig 3C). Across electrodes, the well-trained session showed significantly higher power compared to novel1 ($t_{4812} = 8.451$ $p < 0.001$), while power during the novel1 and novel2 sessions did not significantly differ ($t_{4812} = 1.470$ $p = 0.14168$, *corrected p-value* = 0.010). It is important to note that we chose an *a priori* time window (-250 - 0 msec prior to port-withdrawal) to quantify this performance effect. We also used a data-driven approach to examine whether other epochs exist during which 20-40 Hz dynamics statistically vary with performance level and along the CA1 axis (Fig 3B).

Lastly, we tested whether the observed performance-related effects are specific to 20-40 Hz dynamics or generalize to other bands, including theta (4-12 Hz), the slow gamma range non-overlapping with 20-40 Hz (40-55 Hz), and fast gamma (65-100 Hz; Figure 3—figure supplement1). We found that performance-related effects were weaker, and the patterns differed from those of the 20-40 Hz range. Whereas 20-40 Hz power increased with performance, theta and fast gamma power decreased (theta LMEM ANOVA: $F_{2,4806} = 7.681$ $p < 0.001$; fast gamma LMEM ANOVA: $F_{2,4806} = 8.147$ $p < 0.001$) and slow gamma showed a non-linear relationship with session performance (LMEM ANOVA: $F_{2,4815} = 5.993$ $p = 0.003$) This suggest that the 20-40 Hz band is unique in showing a distinctive increase in power with performance level.

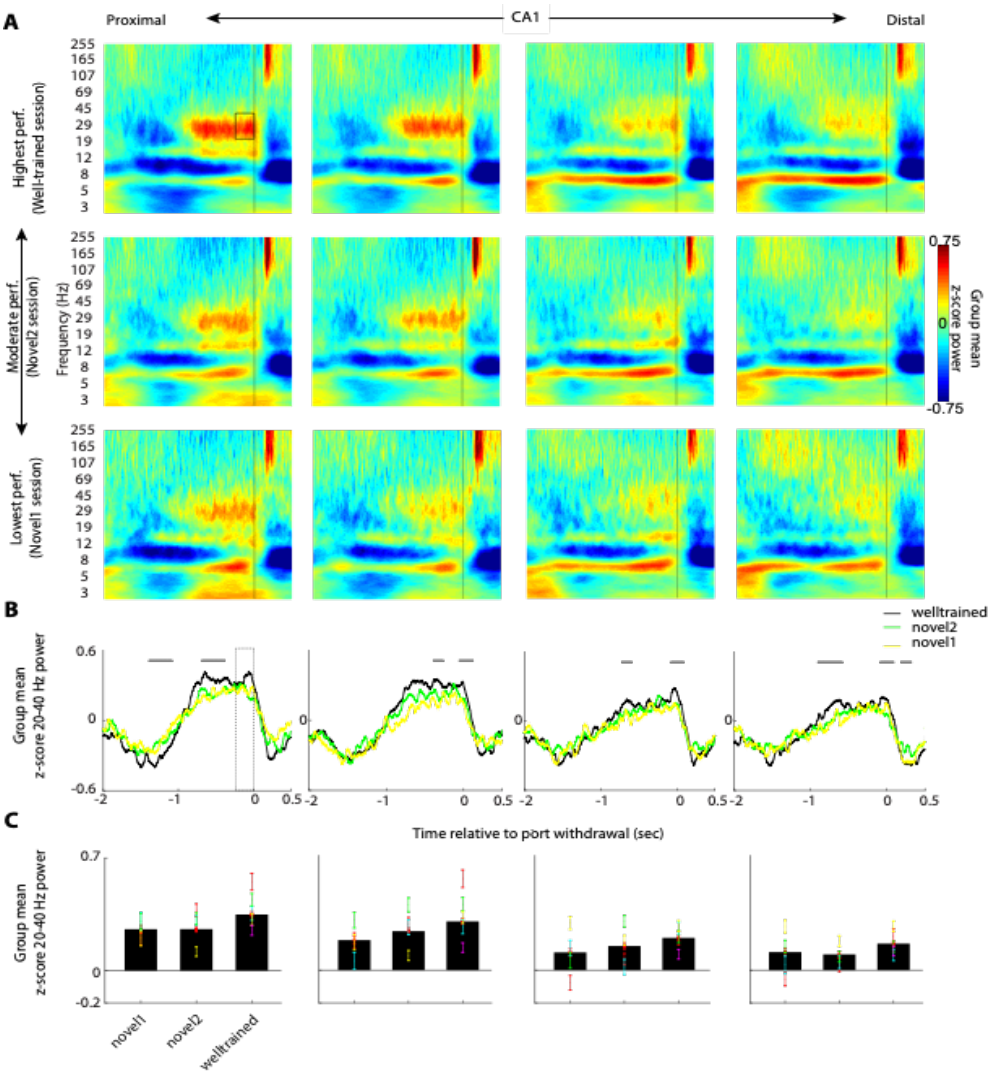


Figure 3. CA1 20-40 Hz power increases with knowledge of the sequence
A. Group peri-event spectrograms across three sessions in which performance levels were low (novel1, first session on novel sequence; *Bottom row*), moderate (Novel2, second session on novel sequence; *Middle row*) and high (well-trained session; *Top*). All plots are aligned to port withdrawal (0ms = port withdrawal) and only include InSeq+ trials. **B.** 20-40 Hz power traces from data in (A). Significant epochs after multiple comparisons correction are indicated with dashed horizontal lines, performed separately for each electrode (cluster-based permutation testing, n=1000 permutations, $p < 0.05$). **C**. Mean z-score power in the 250 ms period prior to port withdrawal (indicated by black and dashed rectangles in A and B, respectively) in each of the three session, shown separately for each CA1 site. Power increased with session performance and in the most proximal sites (significant effect of session $F_{2,4806} = 17.387$, $p < 0.001$, significant effect electrode $F_{3,4806} = 8.9081$, $p < 0.001$). Individual animal values and within animal SEM across trials are overlaid for each barplot. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ indicate significant differences between groups two-tailed t-test, Bonferrnoi correction $p = 0.0167$).

20-40 Hz steady-state power in proximal CA1 is predictive of accuracy for temporal order judgments.

To shed light on the type of processing reflected by 20-40 Hz oscillations, we took advantage of the four different trial types included in our paradigm: InSeq trials that were correctly or incorrectly identified (InSeq+, InSeq-) and OutSeq trials correctly or incorrectly identified (OutSeq+, OutSeq-) (Fig 1C). We began by visualizing power dynamics as a function of the entire time course of trials, focusing on the electrode and session with highest 20-40 Hz power (electrode 1, well-trained session). Aligning data to port entry (Fig 4A, left), we observed that the rising phase (~0-750 ms) was similar across trial types and that differences emerged later in the trial. Specifically, we observed that power was sustained (reaching a steady state) on trials in which animals performed a "hold" response (InSeq+ and OutSeq-), whereas it fell off earlier (during the transient phase) on trials in which the animals performed a "withdraw" response (OutSeq+ and InSeq-). Aligning data to port withdrawal confirmed that the withdrawal response occurs during steady state for the hold responses, but stochastically during the transient response or early steady state for the withdraw-responses (Fig 4A,B, right).

To account for the confounding effect of this *posthoc* observation, we compared power across the four trial types using an early epoch relative to port entry (250-500 ms). This epoch was chosen in aims of minimizing the possibility of capturing modulations in the power dynamics by hold-/withdraw-response behavior. There were no significant differences between trial types in this epoch (LMEM ANOVA: non-significant effect of trial-type $F_{3,749} = 0.480$ $p = 0.693$; Fig 4C, left), suggesting that the 20-40 Hz transient response did not differ with response type (hold/withdraw) nor accuracy (correct/incorrect). Relative to port-withdrawal, and consistent with the posthoc observation, power differed between the hold- and withdraw-response trials in the last 250 msec period (LMEM ANOVA: significant effect of trial-type $F_{3,743} = 9.451$ $p < 0.001$), with significantly higher power during the InSeq+ (hold) compared to the OutSeq+ and InSeq- (withdraw-response) trial types ($t_{743} = -2.829$ $p = 0.005$ and $t_{743} = -4.787$ $p < 0.001$, respectively).

However, power in this epoch did not significantly differ between the hold-response trials (InSeq+ and OutSeq-) (t_{743} = -0.458 p = 0.647).

To more thoroughly examine whether 20-40 Hz power is associated with accurate temporal order judgments, we used a random forest classifier to test whether power in a set of pre-defined epochs and across the CA1 axis predicts response accuracy. Given our post-hoc observation of fundamentally different dynamics for the hold vs. withdraw-response trials, all classification analyses focused on comparing accuracy using the hold-response trial types (InSeq+ vs OutSeq-). This enabled correct vs incorrect comparisons while matching of motor behavior and poke duration distributions (Figure 4—figure supplement1 A). Classification analysis was performed using a 16-D feature vector which included 20-40 Hz power during two epochs relative to port entry (200 - 350, 350 - 500 ms) and two epochs relative to port withdrawal (-400 - -200, -200 - 0 ms) in each of the four CA1 sites (4 epochs X 4 electrodes) (Figure 4—figure supplement1 A, left). We found that these features were not sufficient to discriminate correct and incorrect trials, as classifier performance was at chance levels (mean AUC, TPR and FPR of 0.492, 0.497 and 0.474, respectively) (Figure 4—figure supplement1 A, right). As a follow up, we then used a larger feature space which included time epochs covering the majority of the waveform. We defined a 24-D feature vector which included 20-40 Hz power in four epochs relative to port entry (200-350, 350-500, 500-650 and 650 -1200 msec) and two epochs relative to port withdrawal (-400 - -200 and -200 - 0 msec) in each of the 4 CA1 sites (6 time epochs X 4 electrodes) (Figure 4—figure supplement1 B, left). Using the overlap of lasso (alpha = 0, 0.01, 0.02, and 0.03) and gini impurity based relative importance (Figure 4—figure supplement1 C-D), the 24-D vector was reduced to 3 features most predictive of trial accuracy: 20-40 Hz power during the 650-1200 msec period relative to port entry, in the most proximal sites (electrodes 1-3) (Figure 4—figure supplement1 E).

We found that these three features were sufficient to discriminate between correct and incorrect temporal order judgments (InSeq+ and OutSeq-, respectively). In fact, the classifier

performed with mean areas under the curve (AUC) of 0.588 and 0.477 on the observed and shuffled data, respectively (Fig 4D-E), and a Mann-Whitney U test showed a statistically significant difference between the observed and null AUC values (with higher classification performance on the observed data; U = 839538 $p < 0.001$). Observed compared to null classification (received operating curve) ROC is displayed in Fig 4E, with mean observed classifier AUC, true positive rates (TPR), and false positive rates (FPR) of 0.588, 0.562 and 0.444, respectively.

Posthoc group analyses of the identified feature (650-1200 msec time-epoch) showed a significant effect of trial type (InSeq+ vs OutSeq-; LMEM ANOVA: $F_{1,2406} = 32.687$ $p < 0.001$) and electrode ($F_{3,2406} = 90.573$ $p < 0.001$), but no significant interaction ($F_{3,2403} = 0.193$ $p = 0.901$). Interestingly, power was significantly higher on *incorrect* temporal order judgments (OutSeq- power was higher than InSeq+). Across trial-types, power was significantly higher in electrode 1 compared to electrodes 3 and 4 ($t_{2406} = -12.310$ $p < 0.001$ and $t_{2406} = -13.025$ $p < 0.001$, respectively). Power in electrode 1 was also higher than power in electrode 2, however, this difference did not reach significance ($t_{2406} = -2.215$ $p = 0.027$, corrected p-value = 0.013). Altogether, these results demonstrate that 20-40 Hz steady-state power in proximal CA1 is predictive of accuracy for temporal order judgments during hold response trials, and that incorrect judgments are associated with significantly higher 20-40 Hz power.

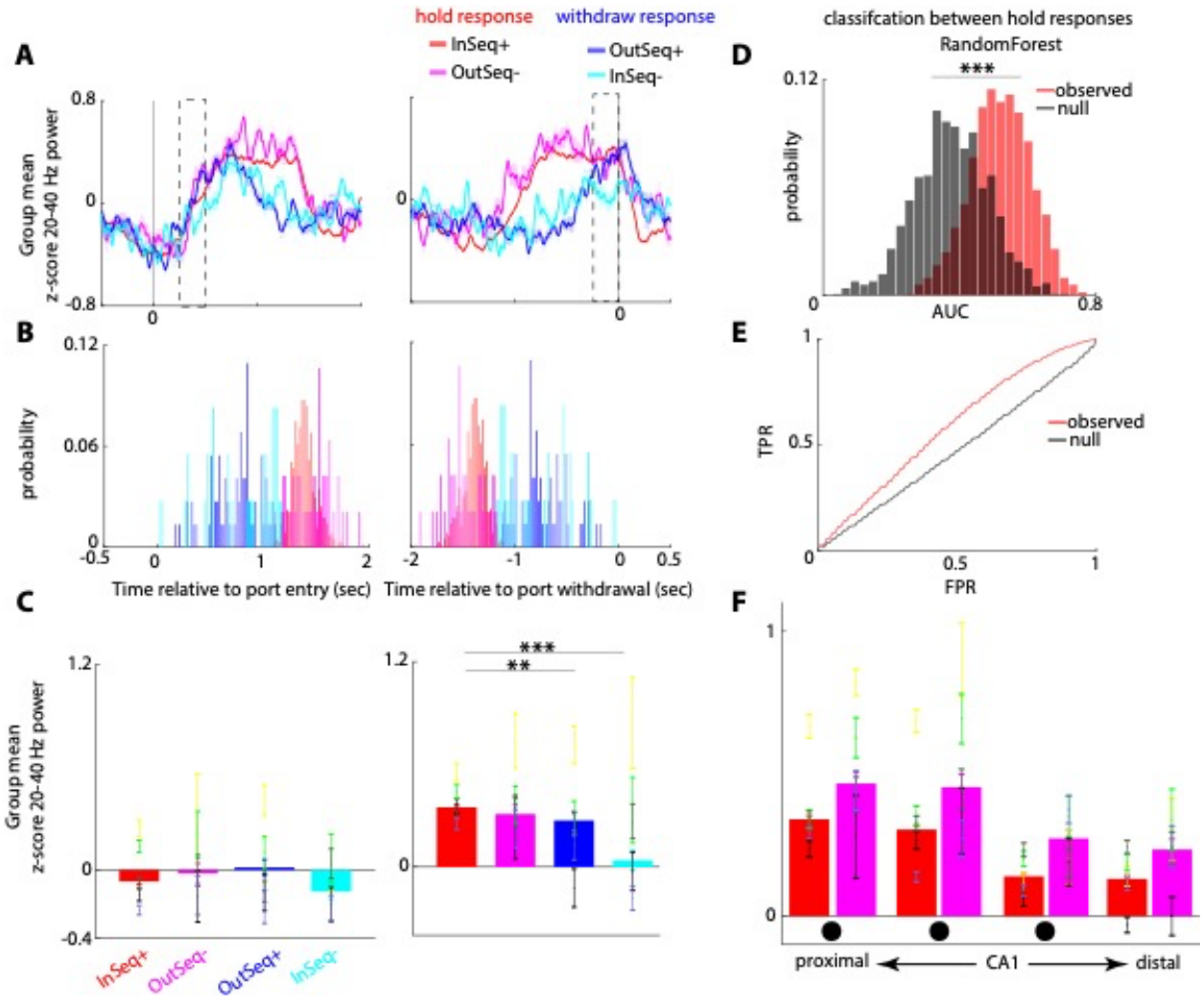


Figure 4. 20-40 Hz power steady state dynamics are disrupted with a withdraw response and predict hold response accuracy

A. 20-40 Hz power traces for four trial types (InSeq+/-, OutSeq+/-) in the most proximal CA1 site, during the well-trained session, with activity locked to both port-entry (left) and port withdrawal (right). **B.** Distributions of time relative to port-entry (left) and port-withdrawal (left), using all trials across all animals used to generate traces in (A). Note that poke durations for hold responses (InSeq+, OutSeq-; shown in red hues) are longer compared to withdraw responses, and 20-40 Hz reached steady-state with longer trial durations. **C.** Mean z-score 20-40 Hz power in the 250 - 500 ms period after port-entry (left) and in the last 250 ms period prior to port withdrawal (epochs indicated by dashed rectangles in A) for each trial type. No differences were observed between conditions in the epoch relative to port-entry (non-significant effect of trial-type $F_{3,749} = 0.4804$, $p = 0.69346$). Power differed between the hold and withdrawal trials types in the 250 ms period relative to port withdrawal (significant effect of trial-type $F_{3,743} = 9.4506$, $p < 0.001$), with significantly lower power during the OutSeq+ and InSeq- (withdraw) trial types compared to InSeq+ (hold) ($t_{743} = -2.8293$ $p = 0.0047902$ and $t_{743} = -4.7869$ $p < 0.001$, respectively). **D.** Random forest classifier mean AUC distributions for classifying between the hold response types (class1: InSeq+, class2: OutSeq-) using observed and null data (mean observed and null AUCs of 0.5879 and 0.4772, respectively). Classifier yielded significantly higher performance when using the observed compared to the null data ($U = 839538$ $p < 0.001$). 3-D classification feature vector

consisted of 20-40 Hz power in the 650-1200 ms period after porty-entry (see supplemental figure 1 for feature selection procedure). **E.** Mean observed and null ROC for the classification in (D). Classifier mean observed AUC, TPR, and FPRs were 0.5879, 0.5619, and 0.4444, respectively. **F.** Mean group z-score 20-40 Hz power in the 650-1200 ms period identified through feature selection (features indicated with black circles), in each of the four CA1 sites. Steady-state power was significantly higher in the OutSeq- compared to the InSeq+ condition (significant effect of trial-type $F_{1,2403} = 8.0819$ $p = 0.0045087$, $t = 5.7172$ $p < 0.001$). Overall power was significantly higher towards proximal CA1 (significant effect of electrode $F_{3,2403} = 81.972$ $p < 0.001$), with electrodes 2, 3, and 4 showing significantly lower power compared to electrode 1 ($t_{2406} = -2.2148$ $p = 0.026869$, $t_{2406} = -12.31$ $p < 0.001$ and $t_{2406} = -13.025$ $p < 0.001$, respectively). Individual animal values and within animal SEM across trials are overlaid for each barplot.

Increased CA1 20-40 Hz power with session performance is selective to hold-response trials

Figure 3 shows increased 20-40 Hz power with session performance and figure 4 shows that 20-40 Hz dynamics predictive of trial accuracy occur late in the trial. In linking these results, we aimed to delineate between two possibilities. The first possibility is that increased power with session performance reflects a direct index of sequence knowledge (i.e., CA1 representation of sequence knowledge or retrieval thereof). This would be supported if increased power with performance would be observed for correct trial types (InSeq+ and OutSeq+). The second possibility is that, given that accuracy predictive dynamics occurred later in the trial (supposedly after memory for order has been retrieved and a post-retrieval decision made), power increases may reflect a post-decision state that varies with session performance (or accuracy). An observation that would support this is that trials reaching a post-decision state (hold responses, InSeq+, OutSeq-) would yield increased power with session performance.

To test these two possibilities, we investigated differences in dynamics across the three sessions separately for the four trial types (Fig 5). Consistent with the latter prediction, using CBPT, we found that power increased with performance level, only for the hold response trial-types (InSeq+, OutSeq-; $p < 0.05$) (Fig 5A). Conversely, power dynamics did not significantly differ with performance level for the withdraw-response trials types (Fig 5B). This included the OutSeq+ trials where order memory retrieval is supposedly intact and improved with training. To

directly compare power in a fixed epoch across all trial types, the OutSeq- epoch was used since it was the overlap between significant epochs for the hold responses. During this epoch, a significant effect of performance level was observed for both hold-response trial-types (InSeq+ and OutSeq-; LMEM ANOVA: significant effect of performance $F_{2,1229,}$ = 14.184 p < 0.001 and $F_{2,173}$ = 12.072 p < 0.001, respectively). For both trial types, epoch power during the well-trained session was significantly higher than that of the novel1 session (t_{173} = 3.984 p< 0.001 and t_{173} = 4.462 p < 0.001, respectively). Lastly, for both trial-types, no significant differences in epoch power between novel1 and 2 sessions were observed (t_{1229} = -0.438 p = 0.662 and t_{173} = 0.748 p = 0.455). Contrarily, epoch power during both withdraw-response trial-types (OutSeq+ and InSeq-) did not significantly differ with performance (LMEM ANOVA: non-significant effect of performance $F_{2,128}$ = 0.384 p = 0.682 and $F_{2,116}$ = 0.796 p = 0.453, respectively). These results support the notion that power increases with session performance may reflect a post-decision state that varies with knowledge of the sequence.

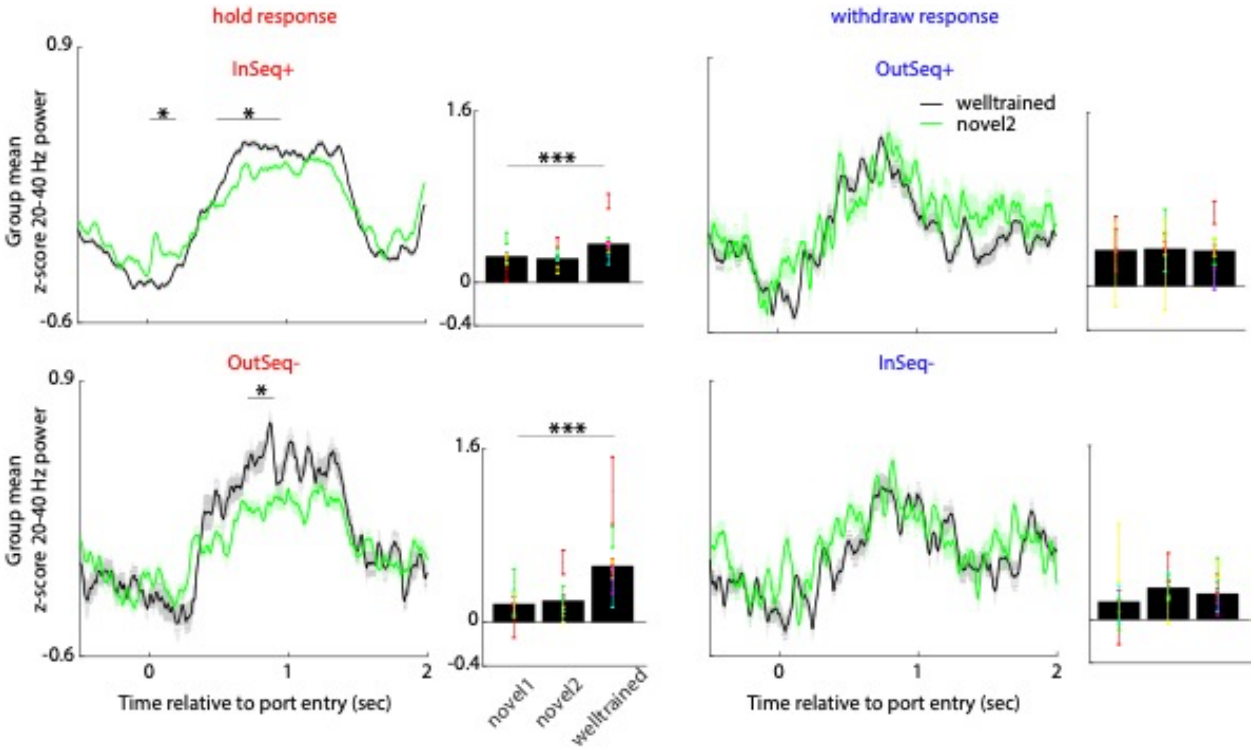


Figure 5. CA1 20-40 Hz power increases with performance level observed only with response maintenance
A. Group mean normalized 20-40 Hz power traces for the hold trial types (InSeq+, OutSeq-) during the well-trained and novel2 sessions using the most proximal CA1 site. Significant epochs after multiple comparisons correction are indicated with dashed horizontal lines, performed separately for each trial type (CBPT, $p < 0.05$). Power was quantified using a fixed epoch across all trial-types (significant OutSeq- epoch; the overlap between hold response identified epochs). Power during the InSeq+ and OutSeq- trial-types (top, bottom, respectively) increased with session performance, with significantly higher power during the welltrained compared to the novel1 session (significant effect of performance $F_{2,1229,} = 14.184$ $p < 0.001$ and $F_{2,173} = 12.072$ $p < 0.001$, respectively) **B.** Same as in (A), but for the withdraw trial types. Power during the OutSeq+ and InSeq- trial-types (top, bottom, respectively) did not significantly differ with performance (non-significant effect of performance $F_{2,128} = 0.38387$ $p = 0.682$ and $F_{2,116} = 0.79624$ $p = 0.45347$). Individual animal values and within animal SEM across trials are overlaid for each barplot. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$ indicate significant differences between groups two-tailed t-test, Bonferrnoi correction $p = 0.025$).

3.4 Discussion

We collected local field potential recordings from the CA1 hippocampal region as rats performed a sequence memory task to identify the frequency content supporting nonspatial information processing. The data presented here expand upon our previous report of 20-40 Hz power during the odor sequence processing periods of the task by evaluating the behavioral relevance and anatomical distribution of this rhythm across the CA1 proximodistal axis, as well as providing a direct comparison with another behavioral state (running). First, we demonstrate that running and odor sequence processing epochs are characterized by different spectral features. Running is associated with increased theta (4-12 Hz) and >24 Hz ranges, whereas odor sequence processing is associated with increased power in theta and 20-40 Hz ranges. Second, we show that in both behavioral states, there are distinct gradients of recruited frequency power along the CA1 proximodistal axis. During odor processing, 20-40 Hz power is higher in proximal CA1, whereas theta is higher in dCA1. During running periods, theta is higher in proximal, whereas >24 Hz power is higher in dCA1. Third, we observed differential performance-related effects across frequency bands; 20-40 Hz dynamics were positively associated with performance across sessions, while theta and gamma demonstrated a negative association. Fourth, we observed differential power dynamics associated with the task's

discriminative response requirements. Both hold-response trials (InSeq+ and OutSeq-) and withdraw-response trials (OutSeq+ and InSeq-) displayed a transient phase where 20-40 Hz power gradually increased at a similar rate. However, dynamics subsequently diverged; 20-40 Hz power declined on withdraw-response (presumed OutSeq decisions) trials while hold-response trials (presumed InSeq decisions) displayed a steady-state period of elevated 20-40 Hz power that remained until the animal withdrew. Fifth, we identified the steady-phase of 20-40 Hz power in proximal CA1 as the key component carrying task-critical information. In fact, this component of the signal (observed on hold-response trials) increased with session performance and was predictive of the accuracy of temporal order judgments. Collectively these results suggest that 20-40 Hz power dynamics in the proximal segment of CA1 are linked to processing the temporal order of nonspatial events and may reflect a post-decision state serving to maintain and protect trial-specific information until a response can be performed.

The nature of this experimental paradigm led to two potential limitations to consider when interpreting the findings. First, the use of a nonspatial discriminative response (hold/withdraw) prevented us from directly equating response duration across all trial types and unexpectedly led to different power dynamics between hold- and withdraw-response trials (possibly attributable to differences in cue utilization across conditions; see Leventhal et al 2012 or merely in trial duration). For that reason, we focused our secondary analyses examining the effects of response accuracy to hold-response trials. Second, to ensure adequate performance, the task requires that the number of OutSeq trials be kept relatively low (otherwise it is unclear which sequence is being tested). This results in an uneven number of observations across trial types which could disproportionally influence a subset of our analyses. However, we controlled for this possibility by conducting pooled analyses using a linear mixed-effects model (LMEM) ANOVA to account for correlations in measures obtained from the same subject, and by matching sample sizes in the classification analyses. Overall, we believe these control analyses

significantly mitigated the potential influence of these confounding factors on the interpretation of our results.

We observed a clear contrast in spectral profiles during spatial compared to nonspatial (odor-sequence) task segments. Consistent with prior reports (Trimper et al., 2017, Chen et al., 2011, Ahmed et al., 2012)., theta and broad-band gamma power were recruited with running. Prior studies suggest that power and synchrony (between distant CA1 regions) shift to higher gamma frequencies with running speed and power correlates with changes in firing rate of individual interneurons. Such shifts are hypothesized to reflect faster transitions between sequential place cell-assemblies (Ahmed et al., 2012). We expand on these observations in demonstrating anatomical theta and gamma CA1 gradients. In contrast to these states and their gradients, during odor processing, 20-40 Hz dynamics emerge and suggest a preferential engagement of proximal CA1 in processing temporal order judgments.

Evidence for functional heterogeneity along the CA1 proximodistal axis has been previously reported (Henriksen et al., 2010; Hartzell et al., 2013; Nakazawa et al., 2016; Ng et al., 2018). Such heterogeneity is not surprising given the known connectivity gradients of the lateral and medial entorhinal cortices (LEC and MEC; Van Strien, Cappaert, and Witter, 2009; Witter et al., 2017), namely that proximal and dCA1 are more anatomically and functionally associated with MEC and LEC, respectively. Consequently, the observation of different proximodistal gradients of oscillatory power across studies may result from differences in task demands, which could promote differential engagement of these entorhinal-CA1 circuits (although other factors, including methodological and analytical approaches, may also contribute to this difference). For instance, Igarashi et al. (2014) showed that 20-40 Hz power and LEC-CA1 coherence were higher in dCA1 during performance of an odor-place association task (whereas we showed higher power in pCA1). Since performance in their task depends on the correct identification of the specific perceptual features that distinguish one odor from another, the power increase in distal CA1 may reflect a stronger engagement of the LEC-dCA1

component of the circuit (as LEC receives strong olfactory input, including direct projections from the olfactory bulb; Haberly and Price, 1978; Agster and Burwell, 2009). In contrast, in our experiment, identification of the presented odor alone is insufficient for correct performance --- the animal must further identify whether the odor is being presented in the appropriate temporal context. The power increase in proximal CA1 we observed may reflect the fact that this additional temporal requirement preferentially engaged the MEC-pCA1 component of the circuit. Interestingly, although MEC is typically associated with spatial navigation functions, this interpretation is supported by a recent study by Robinson and colleagues (2017), in which they demonstrated that optogenetic inactivation of MEC disrupted temporal coding in CA1, while sparing spatial and object coding. Together with the proximodistal pattern we observed, these results suggest that perhaps the MEC→proximal CA1 microcircuit may be important for the processing of nonspatial temporal information.

Another possibility is that the motor output informed by the odor cue is related to the presence of 20-40 Hz (and theta) gradients. In the present study, animals utilized the odor cue to either maintain or terminate a hold posture. Utilization of a sensory cue to inform motor output has been shown to recruit ~20 Hz (beta) in the cortico-basal ganglia (CBG, Leventhal et al., 2012) and the CA1-striatal circuits (Lansink et al., 2017) and in the DG (Rangel et al., 2015), giving rise to a working hypothesis for the functional role of ~20 Hz in sensory-motor integration. Therefore, the observed odor-processing gradients may reflect CA1 involvement with the relevant sensorimotor network. Lastly, the CA1 receives input from a number of other sources including EC, CA2, CA3, and the medial septum (van Strien, Cappaert, and Witter, 2009). Though we previously showed that the spiking activity of a proportion of CA1 neurons is phase-locked to 20-40 Hz oscillations (Allen et al., 2016), the contributions of these upstream structures to the generation of this oscillation in CA1 remain unknown. Importantly, in the present study, we cannot rule out the possibility that volume conduction may contribute to the presence of these rhythms and their gradients. A prior studying performed current source

density analysis for hippocampal 15-35 Hz power, spanning DG, CA3 and CA1, and localized this rhythm to the DG (Rangel et al., 2015). Altogether, future experiments are needed to identify the differential contributions of volume conduction as well as mnemonic and sensorimotor circuits to the observed rhythms and their gradients.

Oscillations in the 20-40 Hz range (typically referred to as beta) occur widely across the brain and have been associated with different functions across brain regions (Engel & Fries, 2010; Schmidt et al., 2019). Of particular interest here is the hypothesis that beta serves a gating function that helps lock in and protect a neural state representing trial-critical information (i.e., maintain the *status quo*; Engel & Fries, 2010). Such a mechanism would be disruptive during encoding or recall processes but advantageous for maintaining a decision until the corresponding response is made, which is consistent with the 20-40 Hz trial-level dynamics we observed in CA1 (low power early in the odor period, but high power in the late period before the response). More precisely, as OutSeq+ trials required animals to perform an “immediate withdraw-response” (i.e., not wait until the signal), we can use the timing of those responses to infer when temporal order decisions occurred. During those trials it was the transient phase of 20-40 Hz power that was consistently observed; any encoding and/or recall, as well as the decision based on such processes, necessarily occurred early in the induction of 20-40 Hz power. This is in line with observations that beta power is attenuated during encoding (Lundqvist et al., 2016; Zavala et al., 2017), and complements findings that transcranial magnetic stimulation at the beta frequency impairs encoding (Hanslmayr et al., 2014). The perspective that beta helps maintain the *status quo* would predict that the steady-state 20-40 Hz period observed on hold-response trials reflects the system actively maintaining an ensemble or network state in support of the decision to perform a hold-response. Though it is possible the post-decision ensemble state reflects a general signal, such as a postural one or a non-specific signal of a temporal context match, our current finding showed that this epoch is predictive of accuracy and our previous work showed that specific odor and temporal context information is

represented at the single-cell (Allen et al., 2016) and ensemble (Shahbaba et al., 2019) levels during this late trial period. It is therefore passible that what is “locked in” during the steady-state beta period is the trial-specific content supporting the animal’s judgment that the odor is presented in the correct temporal context.

This potential gating function of beta is also consistent with the performance-related effects we observed on the session level. Here, we showed that 20-40 Hz power increases with session performance (lowest in novel1, highest in well-trained). High beta during the well-trained session possibly reflects higher levels of gating and effective suppression of distracting processes; this is a high performing session in which the sequence is already well-learned, minimizing the need for additional knowledge acquisition. In contrast, lower beta power in the novel 1 and 2 sessions, which are associated with a higher demand for encoding, may enable flexible knowledge acquisition processes that would otherwise have been interrupted. Consistent with complementary dynamics of beta and gamma rhythms previously reported (Howe et al 2011; Lundqvist et al 2016), we find an inverse pattern for gamma power (and theta) with performance, whereby gamma power is higher during acquisition of a new sequence (novel1-2) and decreases with session performance. This finding highlights the distinction between the functional role of 20-40 Hz and that of the other rhythms, and further corroborates the role of theta and gamma in exploration and encoding processes, respectively.

Interestingly, when considering the relationship between steady-state 20-40 Hz power and performance at both the trial and session levels, we observe a non-linear relationship resembling that of an inverted u-curve. Whereas 20-40 Hz power increased with session performance (from novel1 to well-trained), steady-state power in the well-trained session was higher during OutSeq- (incorrect hold response) compared to InSeq+ (correct hold response). Although this result may seem counterintuitive in light of the proposed gating function of beta, these findings expand on previous work suggesting that low, intermediate, and high beta levels reflect increasing levels of suppression of seemingly distracting processes (Lundqvist et al.,

2016, Schmidt et al., 2019, and Miller et al., 2018). Proximal CA1 higher steady-state beta levels during incorrect trials may reflect over suppression leading to interference with processes needed for successful performance on the task. In other words, the appropriate level of suppression may be advantageous in attenuating distractions, but too high a level may impede essential processes. Consistent with this are clinical reports demonstrating that increased beta in patients with Parkinson's disease is associated with highly sustained, rigid movements (Chen et al., 2007; Pogosyan et al., 2009, Brown, 2007) and difficulty in switching behavioral paradigm (Stoffers et al., 2001); clinically disrupting beta power improves these symptoms, with a side effect of inability to ignore distractions and difficulty in focusing on the task at hand distractions (Cools et al., 2003, Moustafa et al., 2008). Together, these findings point to graded effects of beta, extending to the CA1. Alternatively, higher steady state power may reflect a point of no return, despite retrieval of knowledge that suggests a change in behavior. This latter possibility is consistent with the observation made by Leventhal and colleagues (2012), in which animals were given a second stop cue following an initial go cue. The likelihood of stopping decreased when the second cue occurred during the beta synchronization event elicited from the first cue, reflecting a 'closed gate' on the behavioral state initially cued. Future studies are needed to adjudicate between these possibilities, possibly by specifying the relationship between the information content represented in the ensemble activity and the 20-40 Hz power dynamics.

Evidence from multi-site recordings suggest the 20-40 Hz dynamics we observed in CA1 reflect task-critical, multi-regional integration. Coherence in the beta range between the hippocampus and olfactory regions (Martin et al., 2014; Kay 2014), as well as with the entorhinal cortex (Igarashi et al 2014) have been previously reported during odor dependent tasks. Importantly, these dynamics are not restricted to olfactory processing (visual cues; Lansink et al., 2017, auditory cues; Leventhal et al., 2012) nor do they exclusively coordinate hippocampus with olfactory or medial temporal lobe regions. Lansink and colleagues (2017) demonstrated CA1 beta (15-20 Hz) recruitment as rats responded in the presence of a reward

predictive visual cue that directed navigation (as opposed to goal-directed navigation absent a cue). Notably, during this recruitment neuronal firing was synchronized between CA1 and ventral striatum (Lansink et al., 2017). Specific to our task, structures connected to the hippocampus have also been shown to be critical for performance, including prefrontal, perirhinal, and mid-thalamic regions (Fortin et al., 2016; Jayachandran et al., 2019). As each structure likely contains region-specific representations, beta may serve to bind these regions together to coordinate these distinct task-critical representations and support goal-directed behavior (Fries, 2005). To clarify, multi-regional integration may serve as the general function of beta, which, in our paradigm, is in the service of supporting order memory judgments. Increases in CA1 beta have been reported in other task conditions and demands, such as exploration and novelty (Berke et al 2008, Franca et al 2014), which could reflect multi-regional integration supporting other forms of task-critical information processing. Collectively, these findings provide evidence for the far-reaching role of beta in binding the brain areas and gating their region-specific patterns of activity important for the task at hand.

3.5 Methods

Our group previously published using the same dataset, and a detailed description of the methods can be found in Allen and colleagues (2016). The methods are summarized below.

Subjects

Five male Long-Evans rats were used in this study. Animals were water restricted for optimum task engagement but were provided full access to water on weekends. Proper hydration levels were monitored throughout the experiment. All procedures were conducted in accordance with the guidelines from care and use of laboratory animals published by the National Institutes of Health. All animals were handled according to approved Institutional

Animal Care and Use Committee (IACUC) protocols. Sample sizes were determined using standards in behavioral electrophysiology experiments. Data was recorded from 5 animals (each animal represents several months of work), with each animal providing data from 20 electrodes. See table 1 for trial counts from 1 representative electrode per animal.

Equipment

The apparatus used for this task consisted of a linear track with water ports on either end for water reward delivery. One end of the maze contained an odor port (above the water port) connected to an automated odor delivery system. Photobeam sensors detected when the animal's nose entered and withdrew from the odor port. Detection of nose poke-in triggered odor delivery. Separate tubing lines were used for each odor item, however, all converged at a single channel at the bottom of the odor port. The odor port was kept clear of previous odor traces using a negative pressure vacuum located at the top of the port. A 96-channel Multichannel Acquisition Processor (MAP; Plexon) was used to interface the hardware (Plexon timing boards and National Instruments input/output devices) in real time and record the behavioral and electrophysiological data as well as control the hardware.

Odor sequence task

In this hippocampus-dependent task, rats were presented with a series of five odors delivered one at a time in the same odor port (Fig 1A). In each session, the same sequence was presented multiple times, with approximately half the presentations including all items "in sequence" (InSeq; ABCDE) and the other half including one item "out of sequence" (OutSeq; e.g., AB<u>D</u>DE) (Fig 1B). Each odor presentation was initiated by a nose-poke and rats were required to correctly identify the odor as either InSeq (by holding their nose-poke response until an auditory signal at 1.2 s) or OutSeq (by withdrawing their nose-poke before the signal; <1.2 s)

to receive a water reward (Fig 1C). The presentation of the 5-item sequence terminated earlier if an incorrect response occurred. After termination of a sequence, regardless of performance, the animal was required to run to the opposite side of the track, then return back to the side with the odor port to begin the next sequence presentation. Animals were trained preoperatively on sequence ABCDE (lemon, rum, anise, vanilla, and banana) until they reached asymptotic performance (>80% correct on both InSeq and OutSeq trials; ~6 weeks). Following surgical recovery, electrophysiological data was collected as animals performed the same sequence (ABCDE), followed by two consecutive sessions using a novel sequence (VWXYZ; almond, cinnamon, coconut, peppermint, and strawberry), referred to as novel1 and novel2, respectively (Fig 1D). See Table 1 for individual animal trial counts in each session.

Surgery

Rats received a preoperative injection of the analgesic buprenorphine (0.02 mg/kg, s.c.) ~10 min before induction of anesthesia. General anesthesia was induced using isoflurane (induction: 4%; maintenance: 1–2%) mixed with oxygen (800 ml/min). After being placed in the stereotaxic apparatus, rats were administered glycopyrrolate (0.5 mg/ kg, s.c.) to help prevent respiratory difficulties. A protective ophthalmic ointment was then applied to their eyes and their scalp was locally anesthetized with marcaine (7.5 mg/ml, 0.5 ml, s.c.). Body temperature was monitored and maintained throughout surgery and a Ringer's solution with 5% dextrose was periodically administered to maintain hydration (total volume of 5 ml, s.c.). The skull was exposed following a midline incision and adjustments were made to ensure the skull was level. Six support screws (four titanium, two stainless steel) and a ground screw (stainless steel; positioned over the cerebellum) were anchored to the skull. A piece of skull ~3 mm in diameter (centered on coordinates: -4.0 mm anteroposterior, 3.5 mm mediolateral) was removed over the left hippocampus. Quickly after the dura was carefully removed, the base of the microdrive was

lowered onto the exposed cortex, the cavity was filled with Kwik-Sil (World Precision Instruments), the ground wire was connected and the microdrive was secured to the support skull screws with dental cement. Each tetrode was then advanced ~900 μm into the brain. Finally, the incision was sutured and dressed with Neosporin and rats were returned to a clean cage, where they were monitored until they awoke from anesthesia. One day following surgery, rats were given an analgesic (flunixin, 2.5 mg/kg, s.c.) and Neosporin was reapplied to the incision site.

Electrophysiological recordings

Both spiking and local field potential activity were recorded from the CA1 pyramidal layer of the dorsal hippocampus as rats performed the task (see Allen et al., 2016), but the present study focuses exclusively on a detailed analysis of the LFP activity. Each chronically implanted microdrive contained 20 independently drivable tetrodes, with each tetrode consisting of four twisted nichrome wires (13 μm in diameter; California Fine Wire) gold-plated to achieve a final tip impedance of ~250 kΩ (measured at 1 kHz). Following the surgical recovery period, tetrodes were slowly advanced over a period of ~3 weeks while monitoring electrophysiological signatures of the CA1 pyramidal cell layer (e.g., sharp waves, ripples and theta amplitude). Voltage signals from electrode tips were referenced to a ground screw positioned over the cerebellum. LFP activity was filtered (1.5 - 400 Hz), amplified (1000X), digitized (1 kHz) and recorded to disk with the data acquisition system (MAP, Plexon). Neural activity was first recorded on the odor sequence learned before surgery (ABCDE; “Well-trained” session), followed by two consecutive sessions on the same novel sequence (VWXYZ; Novel1 and Novel2 sessions). At the end of the experiment, recording sites were confirmed by passing current through the electrodes before perfusion (0.9% PBS followed by 4% para- formaldehyde)

to produce small marking lesions, which were subsequently localized on Nissl-stained tissue slices.

Selection of electrodes along the proximodistal axis

In order to sample four representative electrodes along the proximodistal axis of CA1, we chose the first and the last electrodes (most proximal and most distal, respectively) and two electrodes in between which were equidistant. We confirmed the relative spatial distribution of these electrodes as well as their localization within the pyramidal layer of CA1 based on standard spectral properties during baseline, odor sampling and running periods. For each of the four electrodes selected per animal, LFP activity patterns were confirmed in adjacent electrodes (from the remaining subset of 16 electrodes).

Preprocessing and spectral analysis

The raw data was pre-processed using a Butterworth notch filter to remove 60 Hz line noise. Artifact rejection was defined by time indices with voltage values greater or less than 5 standard deviations above or below the mean signal, respectively, of the entire recording in the same channel. Artifact time points were included for the wavelet convolution in order to maintain the temporal structure of the data, but their associated power values were removed before computing the mean and standard deviation for baseline normalization (see below). Any trial containing an artifact was excluded from analyses. For spectral analysis, we utilized the Wavelet toolbox in MATLAB (Mathworks) to generate analytic Morlet wavelets for frequencies between 3 to 250 Hz. These wavelets were tested and verified on a simulated data with known spectral properties. Next, we extracted behavior-locked instantaneous power at the specified frequency ranges. In all analyses, the first trial of each sequence was excluded as it was always preceded by running, whereas the animal was stationary prior to all other trial positions.

Normalization

Instantaneous power is reported as a z-score value relative to the mean and standard deviation of power for a given frequency calculated from a 30-minute subset of the recording from the same electrode. For comparison, we also used two additional normalization approaches. One approach calculated z-scores relative to the other time points within the same trial (0 to 1.5 s for trials aligned to port entry; -1s to 0s for trials aligned to port withdrawal). In the second approach, power for a given time point and frequency within a trial was divided by the sum of the power across all trial time points in the same frequency, which captured percentage increase in power at a given frequency. As all three methods yielded comparable results the reported results relied on the z-normalization to the 30-minute recording subset. As this 30-minute period included a variety of behavioral and cognitive states, including odor sampling, running, grooming, and reward consumption, it offers a better characterization of the variance of spectral dynamics associated with the animal's experiences.

Machine learning analysis

Classification analyses were performed to discriminate between InSeq+ and OutSeq-; these two trial types were matched for response type (hold) but differed in response accuracy (correct/incorrect, +/-). Support vector machine (SVM), logistic regression and random forest were three preselected classifiers which were trained and tested using leave-one-out cross validation, implemented using scikit-learn in python.

For all classification analyses, trial number was fixed between the two classes. Since class 2 (OutSeq-) contained a smaller number of trials compared to class 1 (InSeq+), 1000 iterations were performed where in each iteration classifier performance was computed using a

random subsample of class 1 equal in size to class 2. This yielded an observed distribution of area under the curve (AUC), and true and false positive rates (TPR, FPR) performance values.

Random forest classification performance was tested using the pre-defined time windows (in Fig3-4A,C) with a 16-D feature vector including 20-40 Hz power in four time windows (200-350 and 350-500 msec relative to port-entry and -400 - -200 and -200 - 0 msec relative to port-withdrawal) windows and the 4 CA1 sites (Figure 4—figure supplement1 A). This enabled the investigation of whether the joint values of these features are sufficient to classify performance – an inference that is not extracted from Fig 4C.

Then, implementation was done using a 24-D feature vector defined as 20-40 Hz power in four epochs relative to port-entry (200 - 350, 350 - 500, 500 - 650, 650 - 1200 msec) and two epochs relative to port-withdrawal (-400 - -200, -200 - 0 msec), in each of the four CA1 sites (6 time epochs X 4 CA1 sites) (Figure 4—figure supplement1 B). Random forest showed the best performance and was utilized for all subsequent analyses (see Table 2). Lasso and gini impurity were subsequently used to further reduce the dimension of the starting feature vector, in aims of identifying which of the 24 features contributed to classification accuracy. Using lasso feature selection, sparsity was promoted over a range of alphas (0, 0.01, 0,02 and 0.03) (Figure 4—figure supplement1 C). Features with non-zero weights across all alphas were compared to the five features with highest relative importance using gini impurity (Figure 4—figure supplement1 D). The overlap between lasso feature selection and gini relative importance consisted of 3 features (20-40 Hz power in 650-1200 msec period after port-entry in the three most pCA1 electrodes; Figure 4—figure supplement1 E). These three features were utilized for the final classification analysis presented in Fig 4 D-E.

Replicates

Although it takes several months to train, implant, and record from each animal, the "experiment" focused on three daily sessions per animal (matched across animals). Data from the same animal was not collapsed across sessions. In our design, we view animals as biological replicates and, within each animal, the number of trials as technical replicates. Electrodes can be viewed as biological replicates since we compared effects as a function electrode position within the CA1. Table 1 provides detailed information on the number of trials included from each animal and for each statistical comparison.

Outliers and Inclusion/Exclusion of Data

Standard pre-processing approaches were used to exclude data contaminated by electrical noise or artifacts. Trials with artifacts associated with bumping or touching the headstage (voltage values > 5 SD above the mean) were automatically excluded. Note that this exclusion was performed before (and blinded to) analysis of the results. No additional statistical outliers were removed.

Statistical analyses

Data were first tested for normality using the Anderson-Darling test in Matlab (adtest function). All data met the normality assumption, except for the classification performance output measures (nonparametric tests were performed on the latter). To incorporate all observations in the ANOVAs, while accounting for correlations in scores from the same subjects, a linear mixed-effects model (LMEM) was then used for all group analyses unless otherwise stated. More specifically, the LMEM, which generalizes the conventional general linear model, includes both fixed effects (e.g., electrodes, trials, sessions) and random effects (subjects). Corresponding coefficients and intercepts were estimated for both types of effects, such that the fixed effects reported exclude the variation due to intra-subject correlations. LMEM

fitting and marginal ANOVAs were done using the fitlme and anova matlab functions, respectively. Restricted maximum likelihood method was used for model parameter estimation and cholesky parametrization was used for estimation of the random effects term covariance matrix. It is important to note that standard ANOVAs, either using one value per subject or collapsing observations across subjects, produced a comparable pattern of results, though the LMEM approach is more statistically rigorous. The Bonferroni correction was applied to all subsequent comparisons.

Cluster-based permutation testing was used to agnostically identify group level significant time epochs that differed with session performance, as a function of the full power time course (Fig 3B, Fig 5). Briefly, this involved calculating a t-statistic for each time point, between the novel2 (insufficient correct trials counts to perform this contrast using novel1) and well-trained conditions, generating an observed t-vector. A null distribution of t-vectors was also generated over 1000 unpaired permutations, whereby in each permutation, condition labels were shuffled. A p-value for each time-point was obtained by comparing the observed to the null distribution of t-values at the same time-point, thereby generating a p-vector. To correct for multiple comparisons, clusters of contiguous timepoints with a p <0.05 were identified and compared to the null-distribution cluster size. Observed clusters with sizes larger than the 95^{th} percentile value of null distribution cluster size were considered significant (means of correction for multiple comparisons).

Permutation testing was implemented for the classification analysis, to generate both the observed (see machine learning analysis methods section) and null distributions. Regarding the null distribution, 1000 permutations were done whereby in each permutation, class 1 and 2 observations were randomly shuffled yielding an arbitrary split of 2 classes. Classifier performance was calculated for each permutation, yielding null distributions for AUC, TPR and

FPR performance measures. Lastly, since observed distributions of classifier performance measures were not normally distributed, a Mann-Whitney U test was used to test for differences between the null and observed AUC distributions.

3.6 Supplementary material

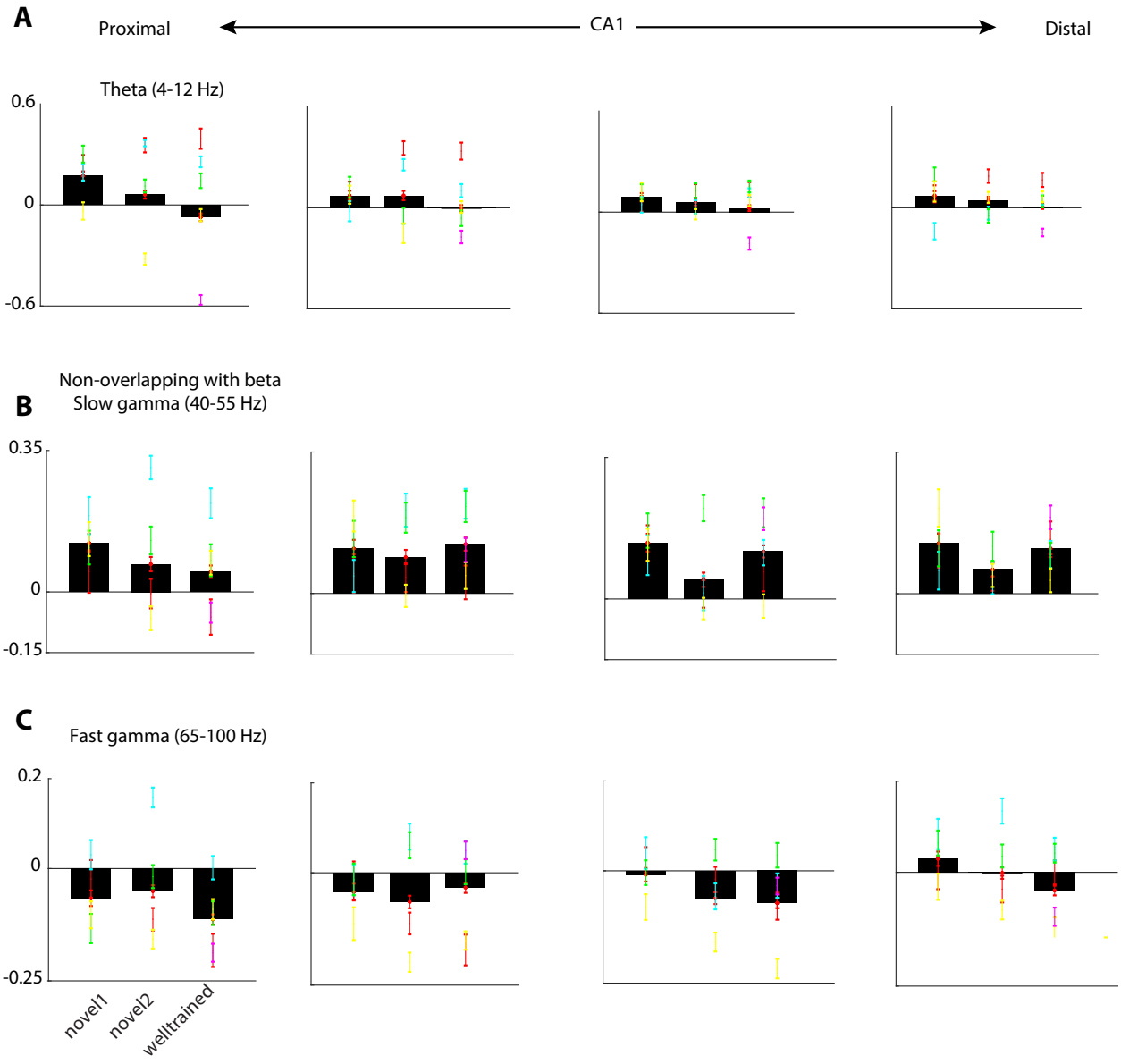
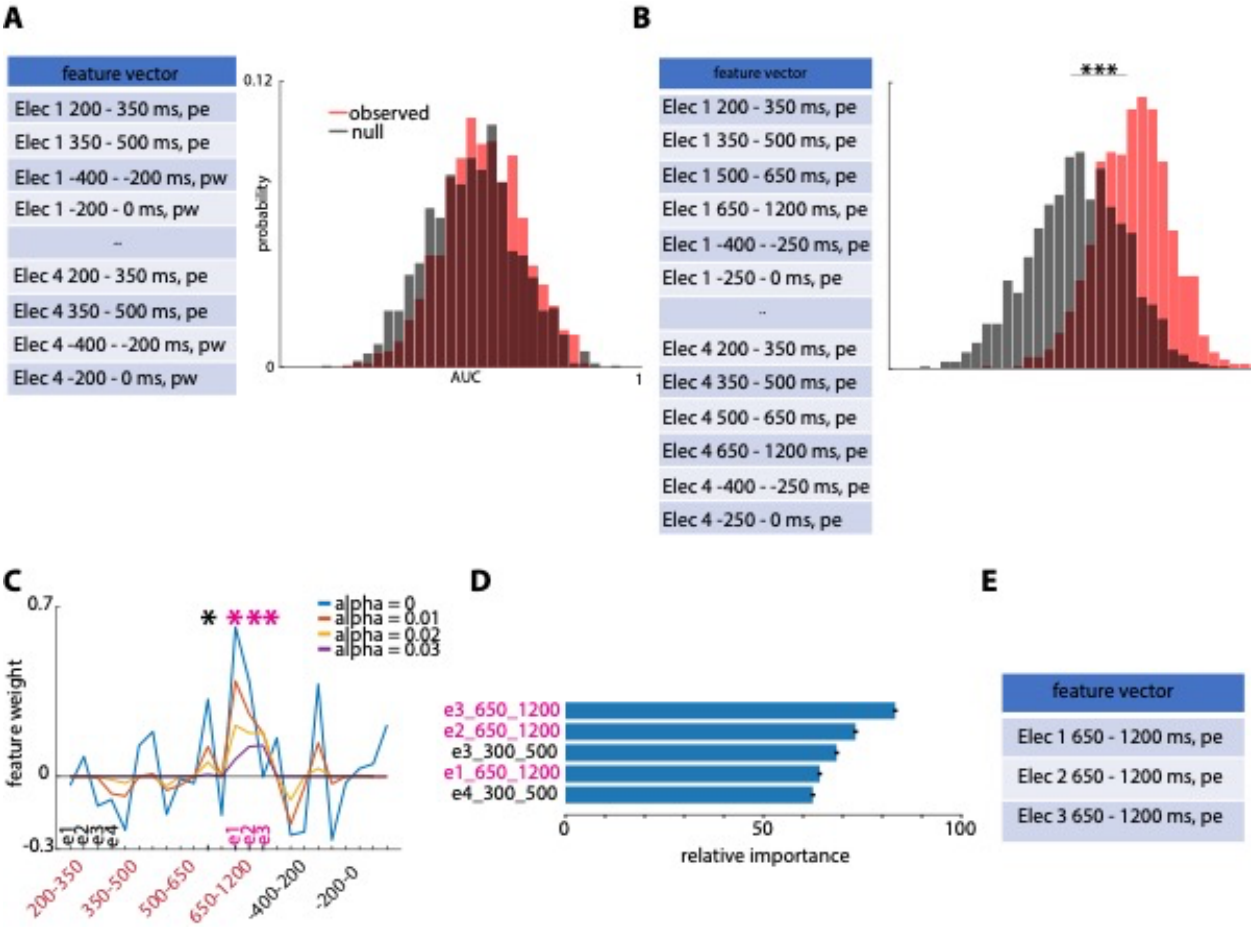


Figure 3—figure supplement1
A-C. Mean z-score theta$_{4\text{-}12\ Hz}$ (A), slow gamma$_{40\text{-}55\ Hz}$ non-overlapping with beta (B) and fast gamma$_{60\text{-}100\ Hz}$ (C) power in the 250 ms period prior to port withdrawal in each of the three session (novel1, novel2, well-trained), shown separately for each CA1 site. **(A)** Theta power significantly decrease with session performance ($F_{2,4806}$ = 7.6814, $p < 0.001$) and along the CA1

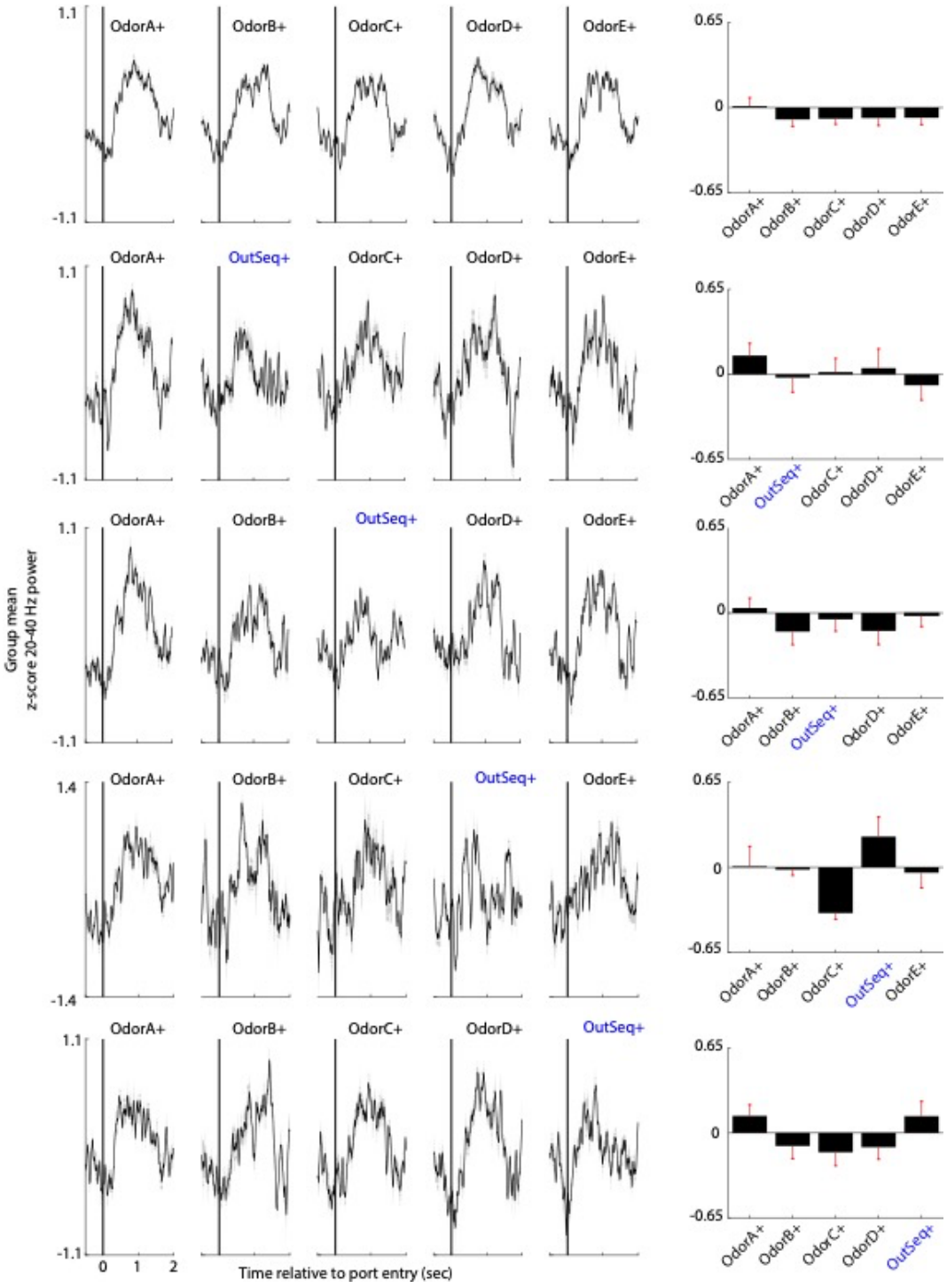
axis (significant effect $F_{3,4806} = 3.5256$, $p = 0.014317$), with a significant interaction between performance and CA1 site ($F_{6,4806} = 4.0794$, $p < 0.001$). **(B)** Slow gamma power showed a significant effect of session performance ($F_{2,4806} = 7.6814$ $p = 0.026347$). Power did not significantly vary along the CA1 axis and the interaction between performance and electrode was non-significant ($F_{3,4806} = 0.25961$ $p = 0.85452$ and $F_{6,4806} = 1.9127$ $p = 0.074972$, respectively). **(C)** Fast gamma power showed a significant effect of session performance ($F_{2,4806} = 8.1471$ $p < 0.001$), CA1 site ($F_{3,4806} = 6.4976$ $p < 0.001$), and a significant interaction between performance and CA1 site ($F_{6,4806} = 3.9286$, $p < 0.001$). Electrode 4 power was significantly higher compared to electrode 1. (Bon ferroni correction: p = 0.01). Individual animal values and within animal SEM across trials are overlaid for each barplot.



Supplementary Figure 2

A. (left) Hypothesis determined 16-D feature vector was defined as 20-40 Hz normalized power in two epochs relative to port-entry (200 - 350, 350 - 500ms) and two epochs relative to port-withdrawal (-400 - -200, -200 - 0 ms) in each of the four CA1 sites. (right) Random forest classifier performance is at chance (mean AUC, TPR and FRP of 0.4923, 0.4965 and 0.4740, respectively) using the feature vector in (A). **B**. Same as in (A), but using data driven definition of a 24-D feature vector including entire power time course, including four epochs relative to port-entry (200 - 350, 350 - 500, 500 - 650, 650 - 1200 ms) and two epochs relative to port-

withdrawal (-400 - -250, -250 - 0 ms) in each of the four CA1 sites. Classifier performed with a mean AUC, TPR and FPR of 0.5727, 0.5532 and 0.4367, respectively. Performance was significantly higher for the observed compared to the null AUC values (U = 821138 $p < 0.001$). **C.** Feature selection from the starting feature vector defined in (B) using lasso at increasing definitions of alpha (promoting sparsity). Features indicated in asterisk contained non-zero weights with all alpha definitions. **D.** Gini impurity-based feature selection displaying the 5 features with highest relative importance out of the feature vector defined in (B). **(E)** Final feature vector for Fig 4 analysis involved the overlapping features identified by the lasso and gini impurity approaches (indicated in magenta in C-D).



Supplementary Figure 3. 20-40 Hz power as a function of position in the sequence
A. (Left) Group mean 20-40 Hz power traces locked relative to port entry, as a function of position in the sequence, for sequences where all items were presented in-sequence and associated with correct performance (InSeq+ -> InSeq+ -> InSeq+ -> InSeq+ -> InSeq+). (Right) Mean power in the 250-500 ms period after port-entry for each item in the sequence displayed

on the left. B. Same as in (A) but for sequence types where all but the second item were presented in-sequence and associated with correct performance (InSeq+ -> OutSeq+ -> InSeq+ -> InSeq+ -> InSeq+). C-E. Same as in (B), but for the OutSeq+ items presented in the 3-5th positions, respectively.

Well-trained

Animal	InSeq+	OutSeq+	InSeq-	OutSeq-
1	135 (55)	25 (24)	5 (2)	5 (2)
2	123 (102)	22 (20)	12 (12)	6 (6)
3	105 (91)	11 (11)	6 (6)	10 (9)
4	135 (131)	27 (27)	14 (13)	17 (14)
5	164 (164)	16 (16)	25 (25)	14 (14)

novel2

Animal	InSeq+	OutSeq+	InSeq-	OutSeq-
1	155 (82)	19 (16)	13 (12)	26 (16)
2	88 (83)	9 (8)	7 (7)	22 (20)
3	136 (86)	3 (3)	11 (10)	15 (12)
4	142 (141)	2 (2)	12 (11)	36 (36)

novel1

Animal	InSeq+	OutSeq+	InSeq-	OutSeq-
1	75 (51)	3 (2)	7 (5)	23 (12)
2	58 (57)	1 (0)	5 (5)	11 (11)
3	92 (75)	0 (0)	8 (7)	11 (9)
4	74 (71)	2 (2)	5 (5)	14 (12)

Table 1.
Trial-type count shown for each session (well-trained, novel1, and novel2) and for each animal. Parenthetical values reflect artifact free (after artifact rejection) trial counts from the most proximal CA1 site per animal.

Data availability

Data are available on Dryad (doi: 10.7280/D11960) and code/scripts used to generate all paper figures and reported statistics are available on Github (https://github.com/FortinLab/Gattas_et_al_2020).

3.7 References

Agster, K. L., & Burwell, R. D. (2009). Cortical efferents of the perirhinal, postrhinal, and entorhinal cortices of the rat. *Hippocampus*, *19*(12), 1159–1186. https://doi.org/10.1002/hipo.20578

Ahmed, O. J., & Mehta, M. R. (2012). Running speed alters the frequency of hippocampal gamma oscillations. *Journal of Neuroscience*, *32*(21), 7373–7383. https://doi.org/10.1523/JNEUROSCI.5110-11.2012

Allen, T. A., & Fortin, N. J. (2013). The evolution of episodic memory. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(SUPPL2), 10379–10386. https://doi.org/10.1073/pnas.1301199110

Allen, T. A., Morris, A. M., Mattfeld, A. T., Stark, C. E. L., & Fortin, N. J. (2014). A Sequence of events model of episodic memory shows parallels in rats and humans. *Hippocampus*, *24*(10), 1178–1188. https://doi.org/10.1002/hipo.22301

Allen, T. A., Salz, D. M., McKenzie, S., & Fortin, N. J. (2016). Nonspatial sequence coding in CA1 neurons. *Journal of Neuroscience*, *36*(5), 1547–1563. https://doi.org/10.1523/JNEUROSCI.2874-15.2016

Axmacher, N., Mormann, F., Fernández, G., Elger, C. E., & Fell, J. (2006). Memory formation by neuronal synchronization. In *Brain Research Reviews* (Vol. 52, Issue 1, pp. 170–182). Brain Res Rev. https://doi.org/10.1016/j.brainresrev.2006.01.007

Bae, G. Y., & Luck, S. J. (2019). Decoding motion direction using the topography of sustained ERPs and alpha oscillations. *NeuroImage*, *184*, 242–255. https://doi.org/10.1016/j.neuroimage.2018.09.029

Bastiaansen, M., Magyari, L., & Hagoort, P. (2010). Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *Journal of Cognitive Neuroscience*, *22*(7), 1333–1347. https://doi.org/10.1162/jocn.2009.21283

Berke, J. D., Hetrick, V., Breck, J., & Greene, R. W. (2008). Transient 23-30 Hz oscillations in mouse hippocampus during exploration of novel environments. *Hippocampus*, *18*(5), 519–529. https://doi.org/10.1002/hipo.20435

Bouthour, W., Mégevand, P., Donoghue, J., Lüscher, C., Birbaumer, N., & Krack, P. (2019). Biomarkers for closed-loop deep brain stimulation in Parkinson disease and beyond. In *Nature Reviews Neurology* (Vol. 15, Issue 6, pp. 343–352). Nature Publishing Group. https://doi.org/10.1038/s41582-019-0166-4

Braver, T. S., Gray, J. R., & Burgess, G. C. (2012). Explaining the Many Varieties of Working Memory Variation: Dual Mechanisms of Cognitive Control. In *Variation in Working Memory*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195168648.003.0004

Brown, P. (2007). Abnormal oscillatory synchronisation in the motor system leads to impaired movement. In *Current Opinion in Neurobiology* (Vol. 17, Issue 6, pp. 656–664). Elsevier Current Trends. https://doi.org/10.1016/j.conb.2007.12.001

Buzsáki, G., & Watson, B. O. (2012). Brain rhythms and neural syntax: Implications for efficient coding of cognitive content and neuropsychiatric disease. *Dialogues in Clinical Neuroscience*, *14*(4), 345–367. www.dialogues-cns.org

Buzsáki, G., & Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. In *Nature Neuroscience* (Vol. 16, Issue 2, pp. 130–138). Nature Publishing Group. https://doi.org/10.1038/nn.3304

Buzsáki, G., & Tingley, D. (2018). Space and Time: The Hippocampus as a Sequence Generator. In *Trends in Cognitive Sciences* (Vol. 22, Issue 10, pp. 853–869). Elsevier Ltd. https://doi.org/10.1016/j.tics.2018.07.006

Chalk, M., Gutkin, B., & Denève, S. (2016). Neural oscillations as a signature of efficient coding in the presence of synaptic delays. *ELife*, *5*. https://doi.org/10.7554/eLife.13824

Chen, C. C., Litvak, V., Gilbertson, T., Kühn, A., Lu, C. S., Lee, S. T., Tsai, C. H., Tisch, S., Limousin, P., Hariz, M., & Brown, P. (2007). Excessive synchronization of basal ganglia neurons at 20 Hz slows movement in Parkinson's disease. *Experimental Neurology*, *205*(1), 214–221. https://doi.org/10.1016/j.expneurol.2007.01.027

Chen, Y., & Huang, X. (2016). Modulation of Alpha and Beta Oscillations during an n-back Task with Varying Temporal Memory Load. *Frontiers in Psychology*, *6*(JAN), 2031. https://doi.org/10.3389/fpsyg.2015.02031

Chen, Z., Resnik, E., McFarland, J. M., Sakmann, B., & Mehta, M. R. (2011). Speed Controls the Amplitude and Timing of the Hippocampal Gamma Rhythm. *PLoS ONE*, *6*(6), e21408. https://doi.org/10.1371/journal.pone.0021408

Colgin, L. L. (2016). Rhythms of the hippocampal network. In *Nature Reviews Neuroscience* (Vol. 17, Issue 4, pp. 239–249). Nature Publishing Group. https://doi.org/10.1038/nrn.2016.21

Colgin, L. L., Denninger, T., Fyhn, M., Hafting, T., Bonnevie, T., Jensen, O., Moser, M. B., & Moser, E. I. (2009). Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature*, *462*(7271), 353–357. https://doi.org/10.1038/nature08573

Cools, R., Barker, R. A., Sahakian, B. J., & Robbins, T. W. (2003). L-Dopa medication remediates cognitive inflexibility, but increases impulsivity in patients with Parkinson's disease. *Neuropsychologia*, *41*(11), 1431–1441. https://doi.org/10.1016/S0028-3932(03)00117-9

Drieu, C., & Zugaro, M. (2019). Hippocampal sequences during exploration: Mechanisms and functions. In *Frontiers in Cellular Neuroscience* (Vol. 13, p. 232). Frontiers Media S.A. https://doi.org/10.3389/fncel.2019.00232

Eichenbaum, H. (2014). Time cells in the hippocampus: A new dimension for mapping memories. In *Nature Reviews Neuroscience* (Vol. 15, Issue 11, pp. 732–744). Nature Publishing Group. https://doi.org/10.1038/nrn3827

Engel, A. K., & Fries, P. (2010). Beta-band oscillations-signalling the status quo? In *Current Opinion in Neurobiology* (Vol. 20, Issue 2, pp. 156–165). Elsevier Ltd. https://doi.org/10.1016/j.conb.2010.02.015

Fortin, N. J., Agster, K. L., & Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience*, *5*(5), 458–462. https://doi.org/10.1038/nn834

Foster, D. J., & Knierim, J. J. (2012). Sequence learning and the role of the hippocampus in rodent navigation. In *Current Opinion in Neurobiology* (Vol. 22, Issue 2, pp. 294–300). NIH Public Access. https://doi.org/10.1016/j.conb.2011.12.005

Haberly, L. B., & Price, J. L. (1978). Association and commissural fiber systems of the olfactory cortex of the rat. I. Systems originating in the piriform cortex and adjacent areas. *Journal of Comparative Neurology*, *178*(4), 711–740. https://doi.org/10.1002/cne.901780408

Hanslmayr, S., Matuschek, J., & Fellner, M. C. (2014). Entrainment of prefrontal beta oscillations induces an endogenous echo and impairs memory formation. *Current Biology*, *24*(8), 904–909. https://doi.org/10.1016/j.cub.2014.03.007

Hartzell, A. L., Burke, S. N., Hoang, L. T., Lister, J. P., Rodriguez, C. N., & Barnes, C. A. (2013). Transcription of the immediate-early gene Arc in CA1 of the hippocampus Reveals activity differences along the proximodistal axis that are Attenuated by advanced age. *Journal of Neuroscience*, *33*(8), 3424–3433. https://doi.org/10.1523/JNEUROSCI.4727-12.2013

Hattori, M., & Sakata, S. (2014). Brain electrophysiological activity correlates with temporal processing in rats. *Behavioural Processes*, *101*, 97–102. https://doi.org/10.1016/j.beproc.2013.09.011

Henriksen, E. J., Colgin, L. L., Barnes, C. A., Witter, M. P., Moser, M. B., & Moser, E. I. (2010). Spatial representation along the proximodistal axis of CA1. *Neuron*, *68*(1), 127–137. https://doi.org/10.1016/j.neuron.2010.08.042

Igarashi, K. M., Ito, H. T., Moser, E. I., & Moser, M. B. (2014). Functional diversity along the transverse axis of hippocampal area CA1. In *FEBS Letters* (Vol. 588, Issue 15, pp. 2470–2476). Elsevier. https://doi.org/10.1016/j.febslet.2014.06.004

Igarashi, K. M., Lu, L., Colgin, L. L., Moser, M. B., & Moser, E. I. (2014). Coordination of entorhinal-hippocampal ensemble activity during associative learning. *Nature*, *510*(7503), 143–147. https://doi.org/10.1038/nature13162

Iversen, J. R., Repp, B. H., & Patel, A. D. (2009). Top-down control of rhythm perception modulates early auditory responses. *Annals of the New York Academy of Sciences*, *1169*, 58–73. https://doi.org/10.1111/j.1749-6632.2009.04579.x

Kay, L. M. (2014). Circuit oscillations in odor perception and memory. In *Progress in Brain Research* (Vol. 208, pp. 223–251). Elsevier B.V. https://doi.org/10.1016/B978-0-444-63350-7.00009-7

Kesner, R. P., Gilbert, P. E., & Barua, L. A. (2002). The role of the hippocampus in memory for the temporal order of a sequence of odors. *Behavioral Neuroscience*, *116*(2), 286–290. https://doi.org/10.1037//0735-7044.116.2.286

Kikumoto, A., & Mayr, U. (2018). Decoding hierarchical control of sequential behavior in oscillatory EEG activity. *ELife*, *7*. https://doi.org/10.7554/eLife.38550

Kilavik, B. E., Zaepffel, M., Brovelli, A., MacKay, W. A., & Riehle, A. (2013). The ups and downs of beta oscillations in sensorimotor cortex. In *Experimental Neurology* (Vol. 245, pp. 15–26). Academic Press. https://doi.org/10.1016/j.expneurol.2012.09.014

Kühn, A. A., Kupsch, A., Schneider, G. H., & Brown, P. (2006). Reduction in subthalamic 8-35 Hz oscillatory activity correlates with clinical improvement in Parkinson's disease. *European Journal of Neuroscience*, *23*(7), 1956–1960. https://doi.org/10.1111/j.1460-9568.2006.04717.x

Kühn, A. A., Kempf, F., Brücke, C., Doyle, L. G., Martinez-Torres, I., Pogosyan, A., Trottenberg, T., Kupsch, A., Schneider, G. H., Hariz, M. I., Vandenberghe, W., Nuttin, B., & Brown, P. (2008). High-frequency stimulation of the subthalamic nucleus suppresses oscillatory β activity in patients with Parkinson's disease in parallel with improvement in motor performance. *Journal of Neuroscience*, *28*(24), 6165–6173. https://doi.org/10.1523/JNEUROSCI.0282-08.2008

Lansink, C. S., Meijer, G. T., Lankelma, J. V., Vinck, M. A., Jackson, J. C., & Pennartz, C. M. A. (2016). Reward expectancy strengthens CA1 theta and beta band synchronization and hippocampal-ventral striatal coupling. *Journal of Neuroscience*, *36*(41), 10598–10610. https://doi.org/10.1523/JNEUROSCI.0682-16.2016

Leventhal, D. K., Gage, G. J., Schmidt, R., Pettibone, J. R., Case, A. C., & Berke, J. D. (2012). Basal ganglia beta oscillations accompany cue utilization. *Neuron*, *73*(3), 523–536. https://doi.org/10.1016/j.neuron.2011.11.032

Levy, R., Ashby, P., Hutchison, W. D., Lang, A. E., Lozano, A. M., & Dostrovsky, J. O. (2002). Dependence of subthalamic nucleus oscillations on movement and dopamine in Parkinson's disease. *Brain*, *125*(6), 1196–1209. https://doi.org/10.1093/brain/awf128

Lisman, J. E. (1999). Relating hippocampal circuitry to function: Recall of memory sequences by reciprocal dentate-CA3 interactions. *Neuron*, *22*(2), 233–242. https://doi.org/10.1016/S0896-6273(00)81085-5

Lisman, J. (2005). The theta/gamma discrete phase code occuring during the hippocampal phase precession may be a more general brain coding scheme. *Hippocampus*, *15*(7), 913–922. https://doi.org/10.1002/hipo.20121

Lundqvist, M., Rose, J., Herman, P., Brincat, S. L. L., Buschman, T. J. J., & Miller, E. K. K. (2016). Gamma and Beta Bursts Underlie Working Memory. *Neuron*, *90*(1), 152–164. https://doi.org/10.1016/j.neuron.2016.02.028

Mably, A. J., & Colgin, L. L. (2018). Gamma oscillations in cognitive disorders. In *Current Opinion in Neurobiology* (Vol. 52, pp. 182–187). Elsevier Ltd. https://doi.org/10.1016/j.conb.2018.07.009

MacDonald, C. J., Lepage, K. Q., Eden, U. T., & Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron*, *71*(4), 737–749. https://doi.org/10.1016/j.neuron.2011.07.012

Mallet, N., Pogosyan, A., Sharott, A., Csicsvari, J., Bolam, J. P., Brown, P., & Magill, P. J. (2008). Disrupted dopamine transmission and the emergence of exaggerated beta oscillations in subthalamic nucleus and cerebral cortex. *Journal of Neuroscience*, *28*(18), 4795–4806. https://doi.org/10.1523/JNEUROSCI.0123-08.2008

Martin, C., Beshel, J., & Kay, L. M. (2007). An olfacto-hippocampal network is dynamically involved in odor-discrimination learning. *Journal of Neurophysiology*, *98*(4), 2196–2205. https://doi.org/10.1152/jn.00524.2007

Mazzetti, C., ter Huurne, N., Buitelaar, J. K., & Jensen, O. (2020). Methylphenidate normalizes aberrant beta oscillations and reduces 1 alpha power during retention in children with ADHD 2 Methylphenidate-induced modulation of brain oscillations. *BioRxiv*, 2020.03.13.990929. https://doi.org/10.1101/2020.03.13.990929

McNaughton, B. L., & Morris, R. G. M. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. In *Trends in Neurosciences* (Vol. 10, Issue 10, pp. 408–415). Elsevier Current Trends. https://doi.org/10.1016/0166-2236(87)90011-7

Miller, E. K., Lundqvist, M., & Bastos, A. M. (2018). Working Memory 2.0. In *Neuron* (Vol. 100, Issue 2, pp. 463–475). Cell Press. https://doi.org/10.1016/j.neuron.2018.09.023

Moustafa, A. A., Sherman, S. J., & Frank, M. J. (2008). A dopaminergic basis for working memory, learning and attentional shifting in Parkinsonism. *Neuropsychologia*, *46*(13), 3144–3156. https://doi.org/10.1016/j.neuropsychologia.2008.07.011

Nakazawa, Y., Pevzner, A., Tanaka, K. Z., & Wiltgen, B. J. (2016). Memory retrieval along the proximodistal axis of CA1. *Hippocampus*, *26*(9), 1140–1148. https://doi.org/10.1002/hipo.22596

Nakazono, T., Sano, T., Takahashi, S., & Sakurai, Y. (2015). Theta oscillation and neuronal activity in rat hippocampus are involved in temporal discrimination of time in seconds. *Frontiers in Systems Neuroscience*, *9*(JUNE). https://doi.org/10.3389/fnsys.2015.00095

Ng, C. W., Elias, G. A., Asem, J. S. A., Allen, T. A., & Fortin, N. J. (2018). Nonspatial sequence coding varies along the CA1 transverse axis. *Behavioural Brain Research*, *354*, 39–47. https://doi.org/10.1016/j.bbr.2017.10.015

O'Keefe J, Nadel L (1978) *The Hippocampus as a Cognitive Map*. Clarendon Press: Oxford.

Pogosyan, A., Gaynor, L. D., Eusebio, A., & Brown, P. (2009). Boosting Cortical Activity at Beta-Band Frequencies Slows Movement in Humans. *Current Biology*, *19*(19), 1637–1641. https://doi.org/10.1016/j.cub.2009.07.074

Rangel, L. M., Chiba, A. A., & Quinn, L. K. (2015). Theta and beta oscillatory dynamics in the dentate gyrus reveal a shift in network processing state during cue encounters. *Frontiers in Systems Neuroscience*, *9*(July), 96. https://doi.org/10.3389/fnsys.2015.00096

Rangel, L. M., Rueckemann, J. W., Riviere, P. D., Keefe, K. R., Porter, B. S., Heimbuch, I. S., Budlong, C. H., & Eichenbaum, H. (2016). Rhythmic coordination of hippocampal neurons during associative memory processing. *ELife*, *5*(JANUARY2016). https://doi.org/10.7554/eLife.09849

Robinson, N. T. M., Priestley, J. B., Rueckemann, J. W., Garcia, A. D., Smeglin, V. A., Marino, F. A., & Eichenbaum, H. (2017). Medial Entorhinal Cortex Selectively Supports Temporal Coding by Hippocampal Neurons. *Neuron*, *94*(3), 677-688.e6. https://doi.org/10.1016/j.neuron.2017.04.003

Schyns, P. G., Thut, G., & Gross, J. (2011). Cracking the Code of Oscillatory Activity. *PLoS Biology*, *9*(5), e1001064. https://doi.org/10.1371/journal.pbio.1001064

Schmidt, R., Herrojo Ruiz, M., Kilavik, B. E., Lundqvist, M., Starr, P. A., & Aron, A. R. (2019). Beta Oscillations in Working Memory, Executive Control of Movement and Thought, and Sensorimotor Function. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *39*(42), 8231–8238. https://doi.org/10.1523/JNEUROSCI.1163-19.2019

Shahbaba, B., Li, L., Agostinelli, F., Saraf, M., Elias, G., Baldi, P., & Fortin, N. (2019). Hippocampal ensembles represent sequential relationships among discrete nonspatial events. *BioRxiv*, 840199. https://doi.org/10.1101/840199

Stoffers, D., Berendse, H. W., Deijen, J. B., & Wolters, E. C. (2001). Motor perseveration is an early sign of Parkinson's disease. *Neurology*, *57*(11), 2111–2113. https://doi.org/10.1212/WNL.57.11.2111

Trimper, J. B., Galloway, C. R., Jones, A. C., Mandi, K., & Manns, J. R. (2017). Gamma Oscillations in Rat Hippocampal Subregions Dentate Gyrus, CA3, CA1, and Subiculum Underlie Associative Memory Encoding. *Cell Reports*, *21*(9), 2419–2432. https://doi.org/10.1016/j.celrep.2017.10.123

Wang, D. D., de Hemptinne, C., Miocinovic, S., Ostrem, J. L., Galifianakis, N. B., Luciano, M. S., & Starr, P. A. (2018). Pallidal deep-brain stimulation disrupts pallidal beta oscillations and coherence with primary motor cortex in Parkinson's disease. *Journal of Neuroscience*, *38*(19), 4556–4568. https://doi.org/10.1523/JNEUROSCI.0431-18.2018

Witter, M. P., Doan, T. P., Jacobsen, B., Nilssen, E. S., & Ohara, S. (2017). Architecture of the entorhinal cortex a review of entorhinal anatomy in rodents with some comparative notes. In *Frontiers in Systems Neuroscience* (Vol. 11, p. 46). Frontiers Media S.A. https://doi.org/10.3389/fnsys.2017.00046

Weiss, S., & Mueller, H. M. (2012). "Too many betas do not spoil the broth": The role of beta brain oscillations in language processing. *Frontiers in Psychology*, *3*(JUN). https://doi.org/10.3389/fpsyg.2012.00201

Wiener, M., Parikh, A., Krakow, A., & Coslett, H. B. (2018). An intrinsic role of beta oscillations in memory for time estimation. *Scientific Reports*, *8*(1), 1–17. https://doi.org/10.1038/s41598-018-26385-6

Wingeier, B., Tcheng, T., Koop, M. M., Hill, B. C., Heit, G., & Bronte-Stewart, H. M. (2006). Intra-operative STN DBS attenuates the prominent beta rhythm in the STN in Parkinson's disease. *Experimental Neurology*, *197*(1), 244–251. https://doi.org/10.1016/j.expneurol.2005.09.016

Zavala, B. A., Jang, A. I., & Zaghloul, K. A. (2017). Human subthalamic nucleus activity during non-motor decision making. *ELife*, *6*. https://doi.org/10.7554/eLife.31007

Chapter 4

Human behavior and associated neural activity revealed through spectral analysis and information theory (deductive)

Theta mediated dynamics of the human hippocampal-neocortical learning systems for memory formation and retrieval

4.1 Abstract

Episodic memory is thought to arise as a function of dynamic interactions between the hippocampus and the neocortex, yet the mechanisms by which these interactions manifest have remained elusive. Here, we report that theta (4-5 Hz) oscillations in humans support directional interactions between the hippocampus and the neocortex when memories are being formed and correctly retrieved. We observe that theta oscillations facilitate information transfer from the hippocampus to the neocortex during learning of new information, and from the neocortex to the hippocampus during the accurate retrieval of previously learned information. These results provide a mechanistic basis for the hypothesized roles of the two learning systems (hippocampus and neocortex) in episodic memory; theta oscillations may facilitate the integration of information into neocortex during encoding and the access of this information from the neocortex during retrieval. Future work can develop theta-based therapeutic interventions for conditions including addiction, post-traumatic stress, and dementia.

4.2 Introduction

The modification of synapses is thought to be necessary for memory formation. Long-term potentiation is a putative basis for this modification that is preferentially synchronized by theta (~5 Hz) oscillations and can be induced by 5 Hz stimulation[1,2]. Observed in live recordings of local

field potentials, the theta rhythm is associated with molecular processes optimal for synaptic plasticity and learning and memory[3–5]. The formation, storage and retrieval of episodic memories are thought to rely on interactions between two learning systems, the hippocampus and the neocortex[6–8]. These interactions theoretically allow for the storage of similar memories using distinct neural codes (i.e. pattern separation) and the recall of previously stored memories from partial or degraded cues (i.e. pattern completion)[9]. Given the overwhelming evidence for the role of theta as a fundamental rhythm involved in memory formation and retrieval[3,10,11], we asked whether hippocampal-neocortical theta oscillatory dynamics support pattern separation. Using human depth electrode recordings during a hippocampus-dependent pattern separation task, we demonstrate a prominent role for theta dynamics. We show that theta power is differentially recruited during successful pattern separation compared to overgeneralization errors, and its phase provides a means of directional information transfer within the hippocampal-neocortical network to facilitate pattern separation. During successful discrimination of similar items (i.e. pattern separation) we observe theta-mediated neocortical→hippocampal directional information transfer. In contrast, during encoding, we observe theta-mediated hippocampal→neocortical information transfer for items later correctly discriminated at retrieval (subsequent memory). Our results identify dynamical systems interactions underlying pattern separation, and more generally, memory formation and retrieval in humans. Additionally, they provide a mechanistic basis for the hypothesized roles of the two learning systems in memory formation and retrieval; theta may facilitate the integration of information into the neocortex during encoding and the access of this information from the neocortex during retrieval. These results address a longstanding question in the neuroscience of memory, paving the path for the development of memory-based therapeutics for various conditions including addiction, post-traumatic stress, and age-related dementia.

4.3 Results

Local field potentials were recorded from eight subjects implanted with concurrent neocortical (NC) and hippocampal (HC) intracranial electrodes during a hippocampus-dependent pattern separation task (Table 1, Fig. 1a). The task consisted of an incidental encoding and a retrieval phase. During encoding, subjects viewed images presented serially and indicated with a button-press whether the item belonged indoors or outdoors. During retrieval, either repeated images from encoding (repeat), similar but not identical ones (lure), or new images (new) were shown, and subjects indicated 'new' or 'old' (Fig. 1a). A 'new' response to a lure reflected successful discrimination between encoding and retrieval images (lure+, lure correct), and an 'old' response indicated failed discrimination (lure-, lure incorrect).

Mean overall task performance was 64.61%, with 82.19% repeat, 92.81% new, and 41.72% lure mean accuracies (Fig. 1b). Neocortical (NC) and hippocampal (HC) cue-responsive electrodes (see methods) were utilized for all analyses (Fig. 1c). NC electrodes included orbitofrontal (OFC), lateral and medial frontal (FRO), temporal (TEMP), cingulate (CING), insular (INS), and entorhinal/perirhinal (EC/PRC) cortices. Those for HC included DG/CA3, CA1, and subiculum. Using serially registered pre- and post-implantation MRI scans, we were able to pinpoint the location of each electrode contact with millimeter precision to achieve subfield-level resolution.

Participant Demographics						Electrode Coverage							
Subject	Age	Sex	Handedness	Years of Education	Seizure onset zone	Orbitofrontal Cortex (OFC)	Frontal Cortex (FRO)	Temporal Cortex (TEMP)	Insular Cortex (INS)	Cingulate Cortex (CING)	Entorhinal/Perirhinal Cortex (EC/PRC)	Hippocampal Dentate Gyrus / CA3	Hippocampus (HC)
1	34	M	A	14	Bilateral HC	4	14	7	1	2	2	3	11
2	32	F	R	12	R lateral TEMP	5	5	4	0	4	5	0	5
3	23	M	R	12	R TEMP	4	18	12	13	6	1	4	9
4	50	F	R	16	Bilateral TEMP	7	7	11	0	6	7	3	13
5	27	M	R	12	Bilateral HC	2	13	11	0	9	0	3	13
6	21	F	R	12	L HC	4	11	11	3	1	3	1	6
7	41	M	R	13	Bilateral lateral TEMP	2	4	17	2	4	2	0	6
8	52	M	R	16	R HC	1	10	13	6	3	0	0	4
Total # of *electrodes*						29	82	86	25	35	20	14	67

Table 1. Study Participants. Participant demographics, clinical notes, and electrode coverage. R: right, L: left.

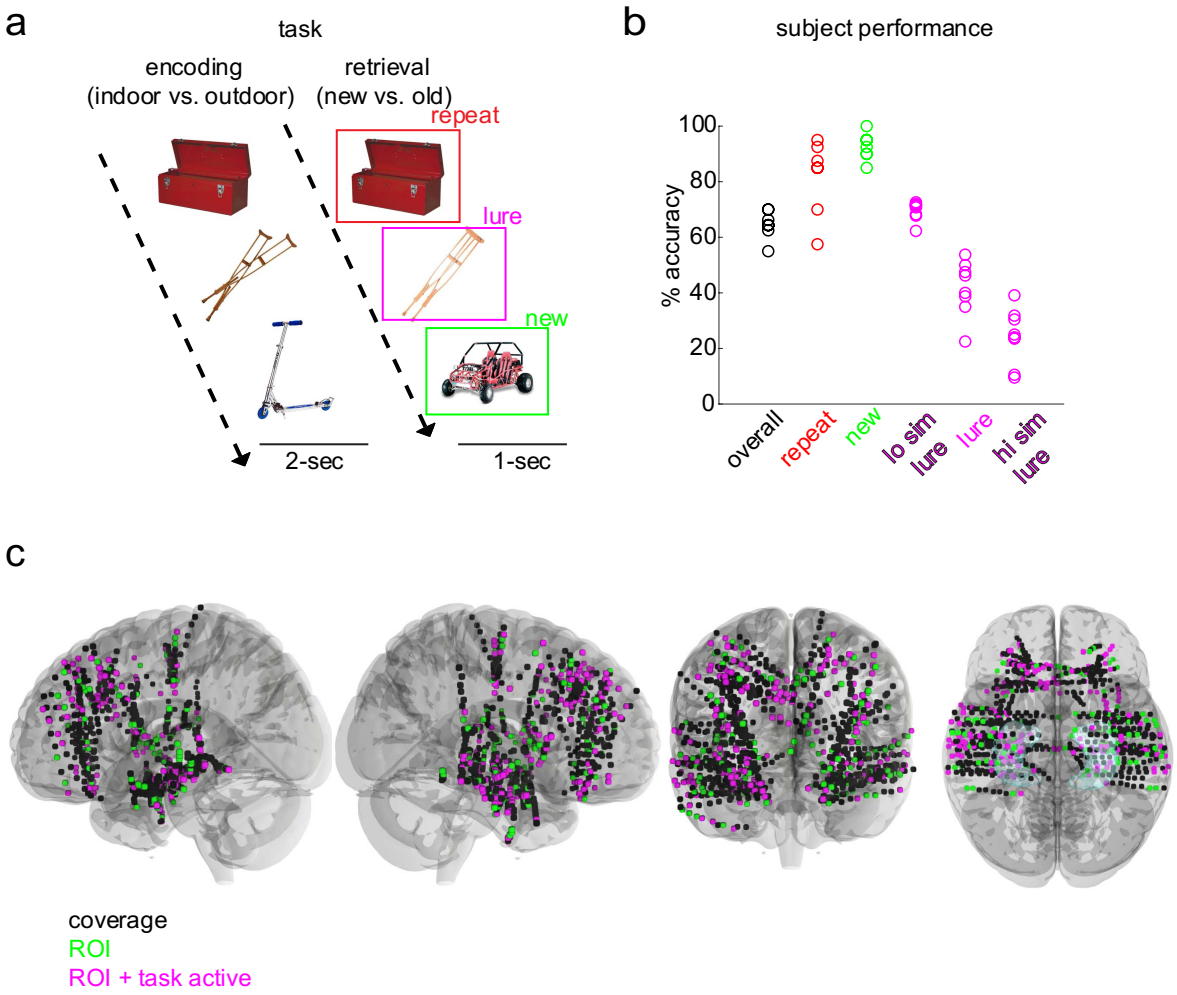


Figure 1. Experimental protocol for investigation of hippocampal and neocortical dynamics

a, Task schematic. The task contained two phases—an encoding and a retrieval phase. In the encoding phase, images of objects were displayed one at a time each for a 2-sec duration. Subjects indicated with a button press whether the images belonged indoors or outdoors. In the retrieval phase, one of three image categories was displayed: an image previously shown in the encoding phase (repeat), a similar but not identical image (lure), or a novel image (new). Following a 1-sec image display, and a 0.5 inter-trial period, a question prompt of 'new or old' appeared to which subjects responded with a button press. **b,** Individual subject behavioral performance on the task (overall accuracy) and specific task conditions (overall, repeat, new, low similarity lures, all lures combined, high similarity lures) **c,** Group coverage displayed in standardized (MNI) brain space. Coverage across all subjects (black), neocortical (OFC, FRO, TEMP, CING, INS, EC) and hippocampal ROIs (green), and the task responsive subset of anatomical regions of interest (magenta) are displayed. A hippocampus region is included in the ventral view.

We first identified the neural population level oscillatory dynamics within the HC that support pattern separation (Fig. 2). Consistent with our prior work[12] and the well-recognized role for theta in memory processing[3,10,11,13], theta power (~4-6 Hz) was significantly higher in the lure+ compared to the lure- condition in the HC (primarily DG/CA3) ($p < 0.05$, paired cluster-based permutation testing) (Fig. 2a). In DG/CA3, only theta survived correction for multiple comparisons (Fig. 2a, bottom panel).

We then quantified whether this theta rhythm is also recruited in the NC. Across all NC areas, a similar narrow-band, 4-5 Hz power was significantly higher in the lure+ compared to the lure- condition ($p < 0.05$, permutation testing) (Fig. 2b). Frontal (~5-6 Hz) and temporal (~ 4-5 Hz) sites were key contributors and showed a statistically significant narrow band theta cluster ($p < 0.05$, permutation testing) (Fig. 2b, Supplementary Figure 1d). OFC, and to a lesser extent INS sites, also recruited slow-frequency oscillations to a greater extent in the lure+ compared to the lure- condition, however in these areas and in CING and EC/PRC, theta clusters did not survive multiple comparisons correction (Supplementary Figure 1). A NC theta discriminatory spatial map using the NC cluster is shown in Fig 2d. Theta power was significantly higher in 30 sites across OFC, FRO, TEMP, CING, and EC/PRC in the lure+ compared to the lure- condition ($p < 0.05$ permutation testing) (Fig. 2d, top panel). We next assessed whether theta was the predominant shared rhythm across these discriminatory sites by generating an unthresholded lure+ vs. lure- difference map. Theta, absent other predominant oscillations, was observed (Fig. 2d, bottom panel). Together, these data suggest that the theta rhythm is preferentially recruited by HC and NC during memory processing supporting pattern separation.

To assess specificity and ensure that this differential recruitment of theta was associated with pattern separation above and beyond what might be expected from accurate memory performance, we compared theta recruitment in lure+ with the new+ and repeat+ conditions (Fid.

2c,d). Hippocampal theta power was significantly higher in the lure+ compared to the repeat+ condition (p = 0.0160, permutation testing) but did not significantly differ from the new+ condition (p = 0.5754). DG/CA3 theta power was higher in lure+ than both repeat+ and new+ conditions and observed on an individual subject basis (Fig. 2c, bottom right panel, Supplementary Figure 1e), however, these differences did not reach significance (lure+ vs. repeat+, p = 0.6853, lure+ vs. new+ p = 0.2338). Regarding NC discriminatory sites, theta power was significantly higher in the lure+ compared to all conditions (lure+ vs. repeat+ p = 0.002, lure+ vs. new+ p = 0.002 permutation testing). Similar condition differences were observed when limiting NC sites to OFC, FRO, TEMP, which were key theta contributors in encoding and retrieval (lure+ vs. repeat+ p = 0.002 and lure+ vs. new+ p = 0.002) (Supplementary Figure 1c), and to lesser extent when agnostically utilizing all NC areas (lure+ vs. repeat+ , p = 0.002 and lure+ vs. new+, p = 0.1319) (Supplementary Figure 1b). Together, these results support the notion that theta recruitment across both DG/CA3 and NC was higher for conditions taxing pattern separation than for standard recognition memory conditions.

Guided by the notion that the longer period of the slow-frequency theta rhythms may facilitate distant interactions limited by conduction velocity across longer synaptic delays[14], we next asked whether theta phase provides a means for behaviorally relevant dynamical interactions between the two learning systems. More specifically, we hypothesized that during retrieval, theta mediates NC → HC information transfer, possibly reflecting the access of memory content from NC stored patterns. To do this, we used an information theory metric, phase transfer entropy (PTE) (see methods), to estimate directional 4-5 Hz information transfer. A PTE value of 0 denotes pure bidirectional interactions, >0 denotes a larger weight to the NC→ HC direction, and <0 denotes a larger weight to the HC→NC direction.

The two systems interacted at retrieval in both directions, with distinct dynamics (Fig. 2e). Consistent with our hypothesis, interactions favored the NC→ HC direction during pattern separation (lure+) (p = 0.0020, permutation testing) (Fig. 2f). This was observed on an individual subject basis (7/8 subjects) (Fig. 2g). Investigating dynamical interactions further as a function of time (PTE-FOT) (Fig 2e) revealed sustained significant differences favoring the NC→HC direction in the 0.6-1 sec post-stimulus period. A repeated measures ANOVA revealed a significant interaction between time and condition, F (50,73400) = 13.679, p < 0.001, and a significant main effect for time on NC-HC interactions F (50,73400) =10.392, p < 0.001. Clear epochs with significant differences between the lure+ and lure- conditions were identified with FDR corrected permutation testing. These dynamical systems interactions were specific to the lure+ accurate discrimination condition in which pattern separation is selectively taxed. Different interactions were observed in the new+ and repeat+ conditions, when examining both the time average and interactions as a function of time (Supplementary Figure 1, f-h).

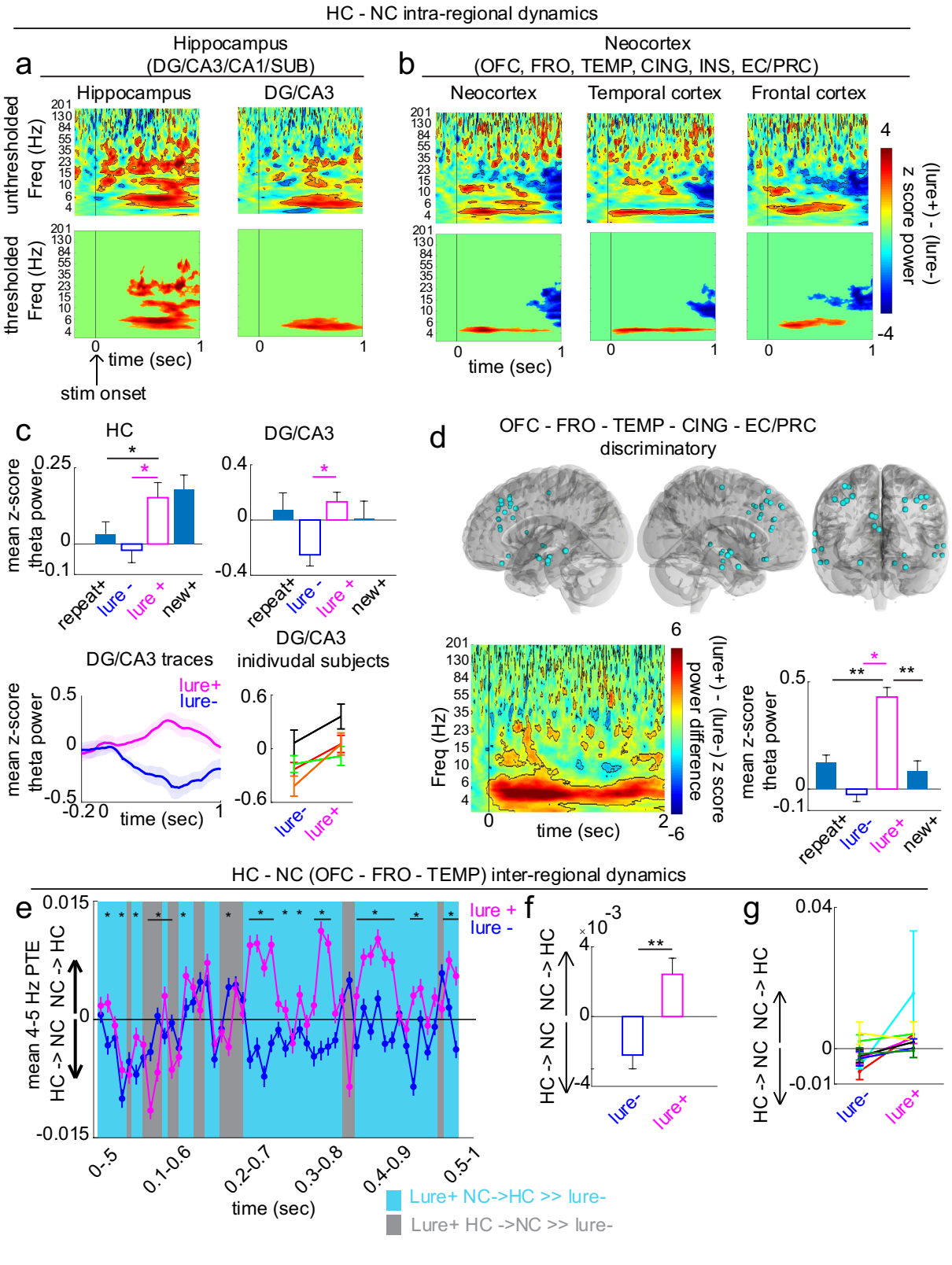

HC - NC intra-regional dynamics
a
Hippocampus
(DG/CA3/CA1/SUB)
Hippocampus
DG/CA3
unthresholded Freq (Hz)
thresholded Freq (Hz)
time (sec)
stim onset
b
Neocortex
(OFC, FRO, TEMP, CING, INS, EC/PRC)
Neocortex
Temporal cortex
Frontal cortex
time (sec)
(lure+) - (lure-) z score power
c
HC
DG/CA3
mean z-score theta power
repeat+
lure -
lure +
new+
DG/CA3 traces
DG/CA3 inidivudal subjects
lure+
lure-
time (sec)
d
OFC - FRO - TEMP - CING - EC/PRC
discriminatory
Freq (Hz)
time (sec)
power difference
(lure+) - (lure-) z score
HC - NC (OFC - FRO - TEMP) inter-regional dynamics
e
lure +
lure -
mean 4-5 Hz PTE
HC -> NC NC -> HC
0-.5
0.1-0.6
0.2-0.7
0.3-0.8
0.4-0.9
0.5-1
time (sec)
Lure+ NC->HC >> lure-
Lure+ HC ->NC >> lure-
f
g

Figure 2. Hippocampal and neocortical intra- and inter-regional dynamics during retrieval

a-b, z-score difference map (lure+ - lure-) in (a) HC and (b) NC areas. Top: unthresholded maps (outlined voxels, p < 0.05). Bottom: significant clusters after multiple comparisons correction (p < 0.05). **c,** Top: mean theta cluster power across 4 condition types (correct (+), incorrect (-)). In HC, theta power was significantly higher in the lure+ compared to the repeat+ condition ($p_{(FDR)}$ threshold = 0.0160). In DG/CA3, no significant differences between conditions, beyond lure+ and lure-, were observed ($p_{(FDR)}$ threshold <0.001). Bottom panel: DG/CA3 theta power traces. Individual subject (n=4) DG/CA3 mean theta power. **d,** Top: spatial map of NC sites with significantly higher NC theta cluster power in the lure+ compared to the lure- condition (p < 0.05). Bottom: difference map and condition specific theta power in NC (cyan) sites. Theta power was significantly higher in the lure+ compared to the repeat+ and new+ conditions ($p_{(FDR)}$ threshold = 0.002). **e,** Phase transfer entropy (PTE) as a function of time (FOT) during the 1-sec post-stimulus period (time x condition interaction p < 0.001, and main effect for time, p < 0.001). PTE significantly differed in particular time epochs between the lure+ and lure- conditions ($p_{(FDR)}$ threshold = 0.019). Blue shading indicates epochs with larger NC→HC weight while gray shading indicates epochs with larger HC→NC weight. **f-g,** Same as (e) but for a (f) single time window (0-1 sec), and (g) in individual subjects. Error shades and bars represent the S.E.M across channels (power), channel pairs (PTE), or trials (individual subjects, 3-D NC map). *p < 0.05, **p < 0.01, ***p < 0.001 indicate significant differences using two-tailed non-parametric paired permutation testing defining channel/channel pairs as observations (1000 permutations).

Next, we investigated the role of theta dynamics during encoding of items that were subsequently discriminated correctly during retrieval (i.e. a subsequent memory analysis). First, we tested whether theta recruitment is larger for items later discriminated (→lure+) compared to those associated with failed discrimination (→lure-). During encoding, 4-5 Hz power was recruited for both stimulus types in HC and NC (Fig 3a). Power was higher in the →lure+ compared to the →lure- condition at slightly different center frequencies (4 Hz for HC and 3 and 5 Hz for NC; Supplementary Figure 2), although these clusters did not survive multiple comparisons correction. In contrast, when limiting NC analysis to the sites discriminating lure+ and lure- trials during retrieval, we observed a statistically significant cluster including the theta range, in which encoding cluster power was significantly higher in items later discriminated (p < 0.05, permutation testing) (Supplementary Figure 2c, traces displayed in Fig. 3b).

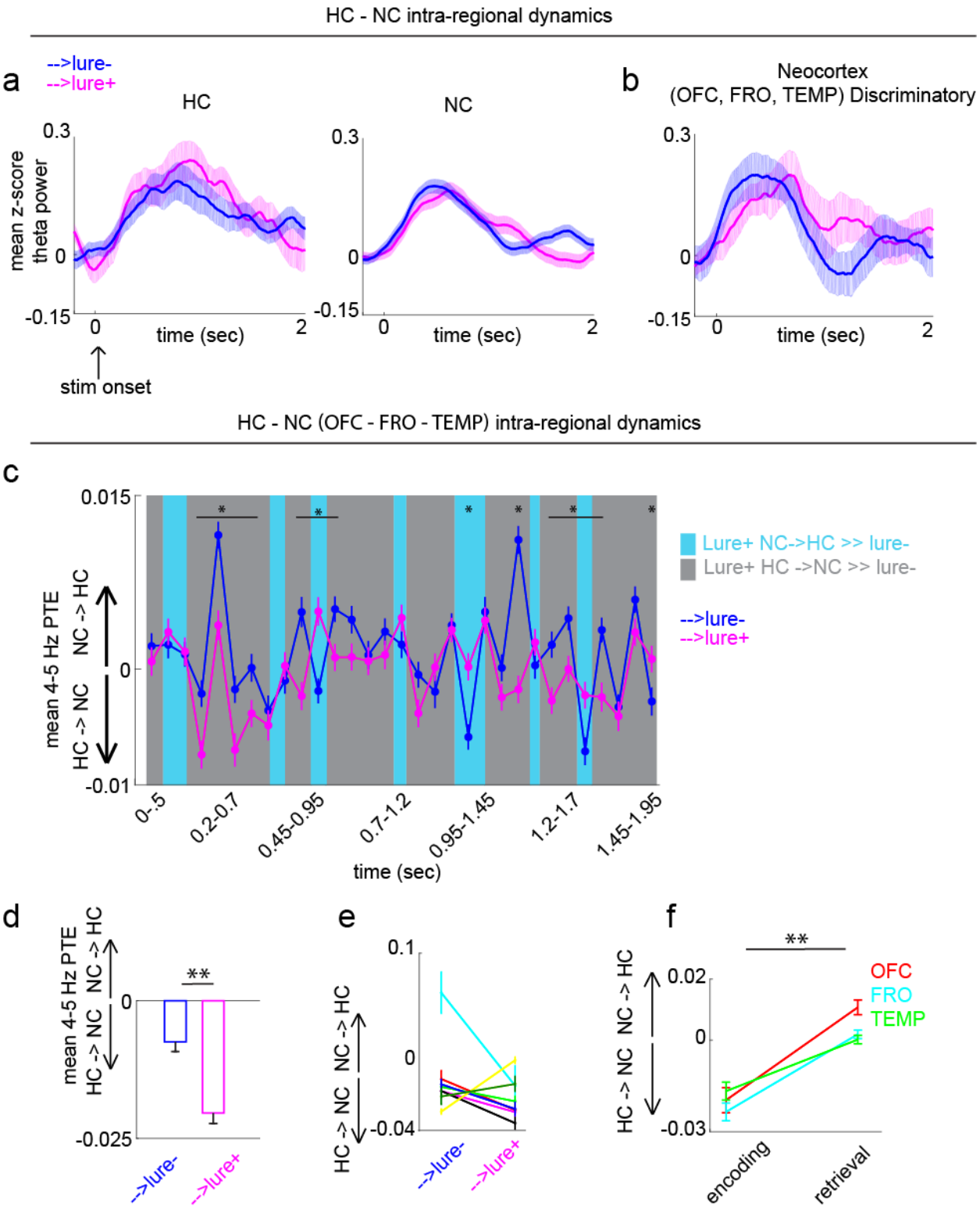


Figure 3. Hippocampal and neocortical intra- and inter-regional dynamics during encoding

a, Encoding HC and NC mean 4-5Hz power time traces for stimuli which were later discriminated (→ lure+) and those associated with failed discrimination (→lure-) in retrieval. **b,** Same as in (a), but limiting NC sites to those showing a lure+ vs. lure- discrimination signal during retrieval. During encoding a statistically significant cluster after multiple comparisons correction (Supplementary Figure 2c) including the 4-5 Hz range was observed in these sites. **c,** PTE as a FOT during the 2-sec post-stimulus onset (time x condition interaction p = 0.010, and main effect for time, p = 0.016). PTE significantly differed in particular time epochs between the lure+ and lure- conditions ($p_{(FDR)}$ threshold = 0.02). Blue shading indicates epochs with larger NC→HC weight while gray shading indicates epochs with larger HC→NC weight. **d-e,** Same as in (c) but for (d) a single time window (0-2 sec), and (f) in individual subjects. **f.** HC-NC PTE values in the →lure+ encoding, and lure+ retrieval conditions, shown separately in each NC site. Error shades and bars represent the S.E.M across channels (power), channel pairs (PTE), or trials (individual subjects). *p < 0.05, **p < 0.01, ***p < 0.001 indicate significant differences using two-tailed non-parametric paired permutation testing defining channel/channel pairs as observations (1000 permutations).

We then asked whether theta-mediated systems interactions during encoding are important for later discrimination. We tested two key hypotheses: 1) Predominantly HC → NC interactions will be observed during encoding, possibly reflecting the integration of HC information into NC patterns 2) A greater degree of HC → NC information transfer during encoding would predict successful discrimination at retrieval. To test the first hypothesis, we compared encoding and retrieval PTE values separately in OFC, FRO and TEMP with the HC. In each region pair, the HC→NC direction was more prevalent during encoding, while the opposite direction of information flow was more prevalent during retrieval (Fig. 3f) (p = 0.002 for each site, permutation testing). To test the second hypothesis, we quantified information transfer in the two condition types over the 2-second post-stimulus period in encoding, and as a function of time in the same period. Encoding interactions were again largely bidirectional, with a significantly larger HC →NC information flow in →lure+ compared to →lure- conditions (p = 0.002) (Fig. 3d). This was observed for most individuals (6/8 subjects) (Fig. 3e). PTE-FOT analysis further identified particular epochs (~0.2-0.7 sec, and 1-2 sec) where the HC→ NC bias was predominant (Fig. 3c). A repeated measures ANOVA revealed a significant interaction between time and condition type, F (30, 44040) = 6.6476, p = 0.010, and a significant main effect for time on NC-HC interactions F (30, 44040) = 5.861, p = 0.016. Once again, clear epochs with significant PTE differences between the lure+ and lure- conditions were identified with FDR-corrected permutation testing. Collectively, the

above results suggest that while systems interactions are largely bidirectional, small biases in directionality during encoding and retrieval may provide a mechanism for the HC's role as an index and the NC's role as the storage site for encoded patterns.

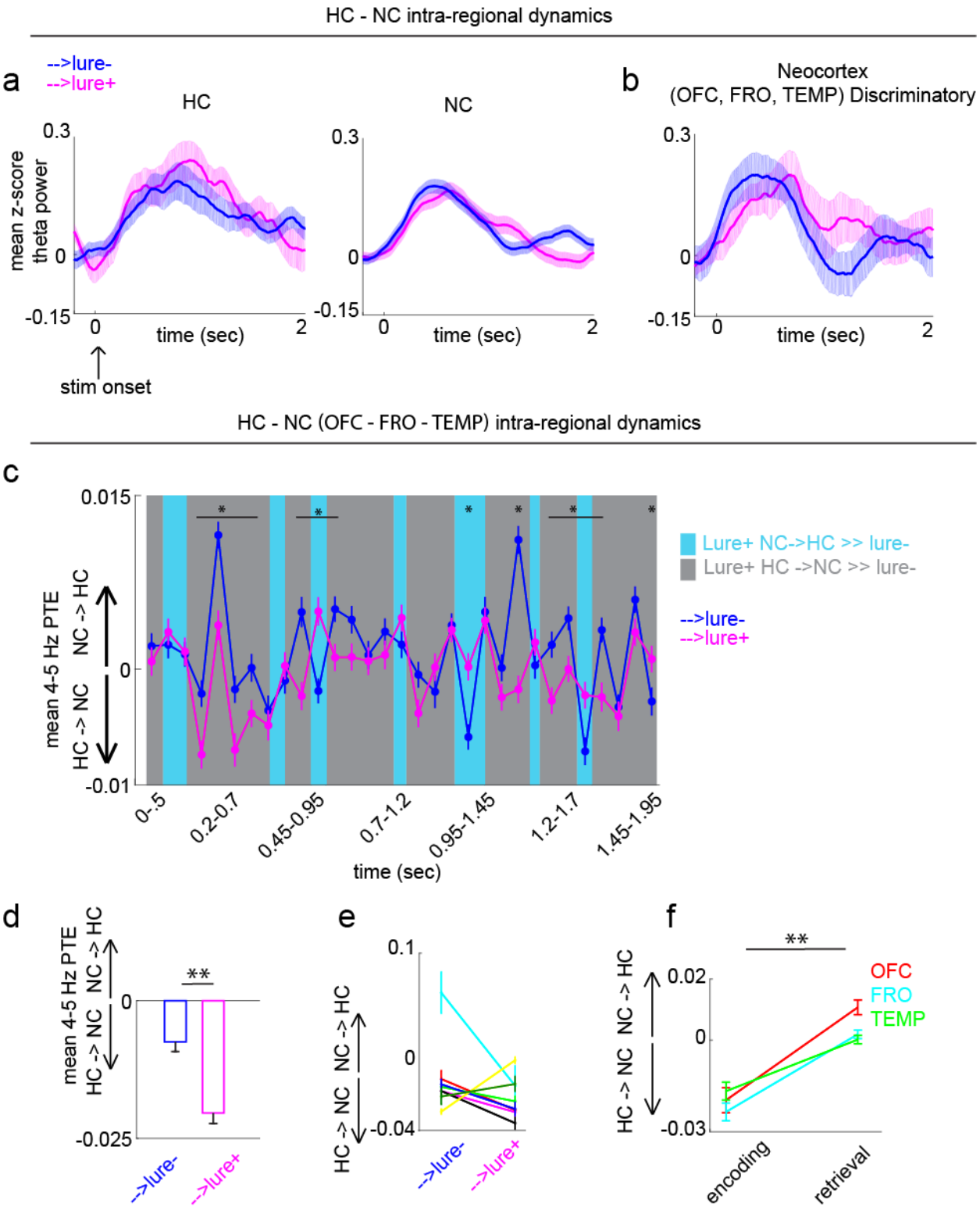


Figure 3. Hippocampal and neocortical intra- and inter-regional dynamics during encoding

a, Encoding HC and NC mean 4-5Hz power time traces for stimuli which were later discriminated (→ lure+) and those associated with failed discrimination (→lure-) in retrieval. **b,** Same as in (a), but limiting NC sites to those showing a lure+ vs. lure- discrimination signal during retrieval. During encoding a statistically significant cluster after multiple comparisons correction (Supplementary Figure 2c) including the 4-5 Hz range was observed in these sites. **c,** PTE as a FOT during the 2-sec post-stimulus onset (time x condition interaction $p = 0.010$, and main effect for time, $p = 0.016$). PTE significantly differed in particular time epochs between the lure+ and lure- conditions ($p_{(FDR)}$ threshold = 0.02). Blue shading indicates epochs with larger NC→HC weight while gray shading indicates epochs with larger HC→NC weight. **d-e,** Same as in (c) but for (d) a single time window (0-2 sec), and (f) in individual subjects. **f.** HC-NC PTE values in the →lure+ encoding, and lure+ retrieval conditions, shown separately in each NC site. Error shades and bars represent the S.E.M across channels (power), channel pairs (PTE), or trials (individual subjects). $*p < 0.05$, $**p < 0.01$, $***p < 0.001$ indicate significant differences using two-tailed non-parametric paired permutation testing defining channel/channel pairs as observations (1000 permutations).

Finally, we investigated whether learning system dynamics supporting early mnemonic discriminatory processes are distinct from the dynamics supporting subsequent response generating mechanisms (e.g. executive function and motor planning) associated with such processes. Within a trial, we refer to the early phase consisting of the 1-sec display of an image as the 'mnemonic epoch', and the later phase consisting of the last 1-sec preceding the subject response (while the question prompt is on the screen) as the 'response generating epoch'. These two epochs were non-overlapping in time. In both systems, a drop in power across a broad range of frequencies (~3-30 Hz) preceding subject response was significantly larger in the lure+ compared to the lure- condition ($p < 0.05$, permutation testing) (Fig. 4a-d). In contrast, this was not statistically observed in the NC discriminatory sites (Fig. 4f). DG/CA3 slow-frequency power drop was also significantly larger in the lure+ compared to the new+ ($p = 0.0240$, paired permutation testing), but not the repeat+ condition ($p = 0.2957$) (Fig. 4g). NC slow-frequency power drop was specific to response judgements associated with successful pattern separation (lure+ vs repeat+ $p = 0.002$, lure+ vs. new+ $p = 0.002$) (Fig. 4h).

We also tested whether this slow-frequency power drop was associated with changes in systems interactions between onset and response periods by estimating the distribution of within-channel onset minus response PTE, separately in each direction (Fig. 4i). With equal probability,

NC-HC pairs are likely to increase or decrease the degree of information transfer upon transition to the pre-response period. This was observed in 4-5 Hz and across 3-20 Hz, and across all NC areas irrespective of discrimination level (Supplementary Figure 3h). Amidst these interaction shifts, the net direction was consistent with a larger NC → HC bias compared to that observed in encoding (p = 0.002 for each site, paired permutation testing) (Fig. 4j). Overall, these results suggest that during the later phase of response preparation and generation (presumably after the mnemonic processing has occurred), there is a drop in theta power in DG/CA3 and NC that was associated with successful pattern separation, as well as an equal probability of an increase or a decrease in information transfer between the two systems. This analysis demonstrates that power dynamics during the response generating epoch are distinct from the mnemonic epoch. This is observed to a much lesser extent when considering phase dynamics, which remain relatively maintained from the mnemonic to the response generating epoch.

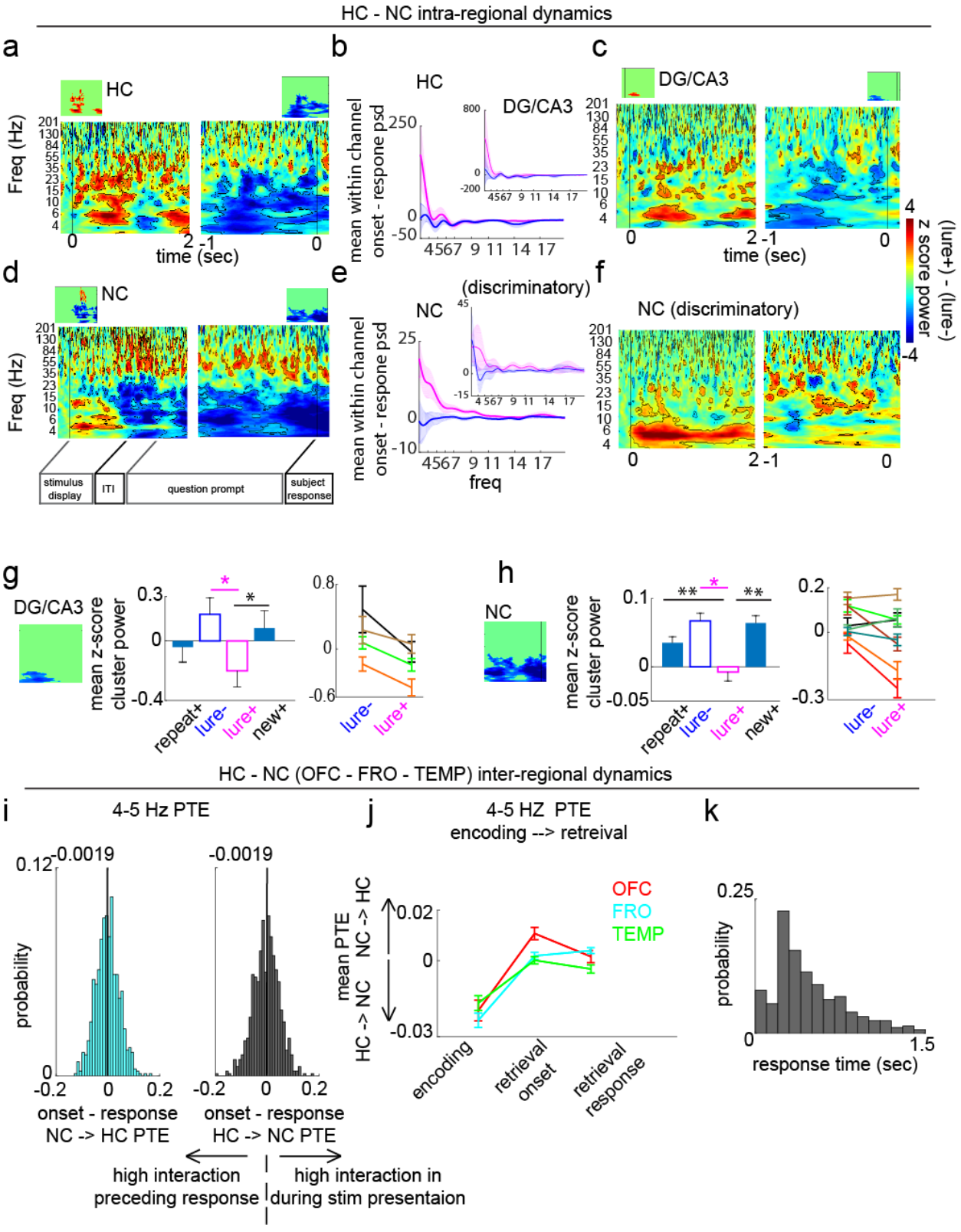

HC - NC intra-regional dynamics
a
HC
Freq (Hz)
time (sec)
b
HC
DG/CA3
mean within channel onset - respone psd
c
DG/CA3
time (sec)
z score power
(lure+) - (lure-)
d
NC
Freq (Hz)
stimulus display
ITI
question prompt
subject response
e
(discriminatory)
NC
mean within channel onset - respone psd
freq
f
NC (discriminatory)
g
DG/CA3
mean z-score cluster power
repeat+
lure-
lure+
new+
h
NC
HC - NC (OFC - FRO - TEMP) inter-regional dynamics
i
4-5 Hz PTE
probability
onset - response
NC -> HC PTE
HC -> NC PTE
high interaction preceding response
high interaction in during stim presentaion
j
4-5 HZ PTE
encoding --> retreival
mean PTE
HC -> NC NC -> HC
OFC
FRO
TEMP
encoding
retrieval onset
retrieval response
k
probability
response time (sec)

Figure 4. Hippocampal and neocortical intra- and inter-regional dynamics during retrieval response period

a,c,d,f, z-score unthresholded and thresholded (insets) difference maps (lure+ - lure-) locked to stimulus onset (left) and response (right) in (a) HC, (c) DG/CA3, (d) NC, and (f) discriminatory NC sites. **b, e,** Mean within-channel subtraction of power spectral density between the 1-sec post stimulus and 1-sec pre-response periods (onset - response PSD) in the (b) HC and NC (e) areas. **g, h,** Mean response locked slow frequency cluster (insets) power in DG/CA3 (g) and NC (h) across 4 conditions on a group, and 2 conditions on an individual subject level. **g,** In DG/CA3, slow frequency power was significantly lower in the lure+ compared to the new+ condition ($p_{(FDR)}$ threshold = 0.0240). **h,** In NC, power was significantly lower in the lure+ compared to the repeat+ and new+ conditions ($p_{(FDR)}$ threshold = 0.0279). **i,** Probability density function of within channel onset minus response PTE (onset - response NC→HC PTE (cyan) and onset - response HC → NC (black)). Median values are indicated on top of each distribution. **j,** PTE values in the →lure+ encoding, and lure+ retrieval onset and response conditions, shown separately in each NC site. **k,** Response time distribution across trial types and subjects. Error shades and bars represent the S.E.M across channels (power), channel pairs (PTE), or trials (individual subjects). *p < 0.05, **p < 0.01, ***p < 0.001 indicate significant differences using two-tailed non-parametric paired permutation testing defining channel/channel pairs as observations (1000 permutations).

4.4 Discussion

In summary, we have shown that theta power in the HC (particularly DG/CA3) and NC supports pattern separation, and that theta phase provides a potential mechanism for directional information transfer (measured by phase transfer entropy) in the HC→NC direction during encoding and in the NC→HC direction during retrieval. These early increases in theta power during the mnemonic epoch are followed by a pre-response drop, however, information flow distributions appear to be largely sustained. Overall, these results provide a putative mechanism for hippocampal-neocortical information exchange related to encoding and retrieval of episodic memories that is consistent with the hippocampus' role as an indexing device.

Prior studies investigating NC and HC involvement in humans mainly used methods from which directionality cannot be inferred[13], focused primarily on sleep[8], did not concurrently record from HC and NC[15], or were limited in neocortical coverage[16]. Here, we identify the dynamics of oscillatory power and phase-dependent directional interactions supporting memory formation (new+) and retrieval (repeat+), as well as formation and retrieval of events underlying successful

pattern separation (lure+) vs. overgeneralization errors (lure-). Each of these cognitive states is supported by unique interactive dynamics between the NC and HC. These results provide empirical evidence for the role of theta and theta-mediated interactions in supporting pattern separation, and more generally memory formation and retrieval. They form a mechanistic basis for the long-hypothesized role of theta in episodic memory processing, directional interactions hypothesized by the memory indexing theory[6,17] and computational models based on the two learning systems[7,18–21].

It is critical to note that bidirectional interactions between HC and NC have been recently investigated in human physiology recordings[16]. In this study, the authors report that decreased NC alpha/beta (8-20 Hz) power precedes and predicts increased HC fast gamma (60-80 Hz) during episodic memory encoding, whereas increased HC slow gamma (40-50 Hz) power precedes and predicts decreased NC alpha/beta power during retrieval. While at first our results may appear at odds with these findings, there are key differences across studies that are important to note. Similar to their findings, we qualitatively observed decreased NC alpha (~7-15 Hz) during encoding for subsequently discriminated items (Fig S2a,d). We also observed increased HC slow gamma (~35-55 Hz, Fig 2a, 4a) and decreased NC alpha/beta (~12-23 Hz, Fig 2b) for successfully discriminated items during retrieval. However, we chose to focus *a priori* on the theta rhythm given its known role in long-range communications. In our results, it also was a shared signal across the NC and HC (in particular DG/CA3), which makes it a likely candidate for information transfer. We used an information theoretic approach (phase transfer entropy) to determine whether there was meaningful information transfer across regions. The latter cannot be inferred purely from correlational data. Finally, we demonstrate temporal dynamics of information exchange that suggest a more complex interplay among the two regions during both encoding and retrieval. Thus, our reported results complement previously published work and

additionally provide a putative mechanism by which indexing can occur in interleaved and directionally biased HC-NC interactions.

Since the 1930s, the theta rhythm has been observed across species and behavioral states[10,22]. Amongst a diverse set of its behavioral correlates, a subset includes arousal, attention, decision making, learning and memory, working memory and motivation[10,13]. These processes all require the processing of external inputs[10]. The hippocampal theta rhythm is thought to function as a temporal organizer, enabling first and higher order temporal linkages amongst cell assemblies and converting such orders to synaptic strengths[10]. This is supported by experimental work showing that after controlling for other factors, theta was primarily associated with temporal ordering of experiences[13,23,24]. Theta cycles may separate encoding and retrieval epochs by alternating processing modes during the peak and trough [11]. During the peak, LTP occurs and more predominant inputs from EC to CA1 prevail, possibly facilitating encoding, while LTD occurs during the trough with predominant CA3 inputs to CA1, possibly facilitating retrieval[25,26]. Evidence for theta's role has been demonstrated in its temporal organization of place[27] and grid[28] cell firing (phase precession)[29], and cell encoding time intervals[30]. These findings point to theta as a potential mechanism to encode space and time, the basic ingredients for episodic memory, thereby serving as a fundamental neural code in memory processing.

With respect to learning systems interactions, theta is proposed to encode and temporally bind representations across neocortical modules, as well as mediate the interactions between HC and NC with the HC acting as an index generator[3]. Our results could be explained by this model, providing a mechanistic basis for memory indexing theory. The involvement of the two learning systems in this hippocampus-dependent pattern separation task is predicted by indexing theory and is demonstrated by theta recruitment in HC and NC sites during both encoding and retrieval.

According to the indexing theory, during encoding, a unique spatiotemporal neocortical pattern is generated upon an experience, which is strengthened by the hippocampus[6,17]. This strengthening could be driven by modifications of intracortical and cortical-hippocampal synapses when inputs arrive at the peak of theta. It is therefore possible that the observed increase in NC encoding theta power and the theta-phase mediated HC→NC information flow reflects neocortical learning, strengthened by the hippocampus. The complementary learning systems theory has been recently updated to demonstrate that the neocortex is capable of such rapid learning if new information is consistent with existing schemas[7,31], which is supported by experimental evidence[32,33]. Given the schema-consistent objects viewed in this study, it is possible that integration of information into neocortical connections occurs rapidly during the encoding phase.

Indexing theory also proposes that upon encountering a partial cue, neocortical modules can be activated and accessed by the hippocampus to facilitate retrieval. Consistent with this prediction, we observe that the increase in NC theta power precedes that of the HC during successful retrieval and information flow begins to predominantly favor the NC→HC direction. These two observations could reflect reactivation of a neocortical pattern and its access by the hippocampus to enable comparison of the similar lure to an existing memory, a necessary process for accurate discrimination (i.e. recall-to-reject)[20,34]. In line with the notion that theta power likely supports the aforementioned early mnemonic processes, is that theta power as well as surrounding slow frequencies are significantly diminished preceding response, and half of the NC-HC electrode pairs show a decrease in information transfer. This, in conjunction with prior work showing increased hippocampal and DLPFC gamma power in the same task preceding response[35], suggest relatively more local neocortical and hippocampal activity during the decision phase.

A possible alternative explanation for theta power dynamics during successful encoding and retrieval is active sensing and exploration that facilitates memory encoding[10,22,36–39]. This is a possible explanation for neocortical theta increases during encoding of items that are subsequently discriminated. However, this is unlikely to be driving condition differences during retrieval since the new+ condition is more novel compared to lure+ and does *not* show similar increases in theta power or interaction dynamics. Another possibility is that theta facilitates retrieval of the context in which memory of the item was generated rather than memory of the item itself[40,41]. However, this again would not explain the differences between lure+ and repeat+, both of which presumably have accurate stimulus-specific context retrieval that leads to accurate recall. Finally, an intriguing possibility is that oscillatory inhibition varies as a function of theta phase, with increasing levels of inhibition promoting the suppression of competing memory traces to facilitate accurate retrieval[13,42]. This could be a mechanism by which encoding and retrieval dynamics contribute to successful discrimination among similar items, although this possibility needs to be directly tested by manipulating inhibition.

Although a preponderance of evidence exists for the role of theta in episodic memory processing, such a role may be much more general. For example, theta could be a mechanism for synaptic modification with experience and for routing and integration[22] of multimodal information between behaviorally recruited circuits. This could explain the observation of theta across numerous cognitive states and its enhanced recruitment as a function of increased sensory and mnemonic processing[10,13]. Theta's long period theoretically increases its coding capacity and makes it advantageous for the routing of information within and across distributed networks[15] overcoming the constraints of conduction velocity and longer synaptic delays[3,14]. This is further supported by the notion that the leading cellular contributors to theta generation in the hippocampus, GABAergic interneurons[43,44], are highly connected, long-range projecting hub cells orchestrating network synchrony[45]. Additionally, individuals with higher PFC-HC white matter

integrity tend to have high slow-frequency power recruitment (delta/theta), and better long-term memory[46,47]. Finally, intracranial theta burst stimulation of the entorhinal region, where hippocampal/neocortical exchange takes place, has been previously reported to improve memory function in patients with epilepsy[46,47]. Altogether, theta may be a key contributor to routing inputs and integration of multimodal information into synapses —this would manifest across cognitive domains including, but not limited to episodic memory.

Despite the prominent role of theta oscillations in primate and rodent studies, it is important to note that not all species utilize theta to temporally organize information in episodic memories. For example, the hippocampal formation in bats exhibits temporal coding and synchronization in the absence of low-frequency oscillations, suggesting that rhythmicity is not necessary for effective phase coding[48,49]. From an evolutionary standpoint, theta rhythmicity could be one of several possible mechanisms that enable memory formation and retrieval.

In conclusion, our results provide a novel mechanistic account for indexing theory by demonstrating a role for theta-mediated intra and inter-regional dynamics during encoding and retrieval to support accurate memory discrimination. Future studies are needed to identify the molecular and cellular processes underlying the observed role of theta and whether theta codes for memory representations directly or facilitates their exchange. Ultimately, identification of episodic memory system dynamics can lend to developing memory-based therapeutics for various conditions including addiction, post-traumatic stress, and age-related dementia.

4.5 Methods

Experiment paradigm

Electrophysiological recordings were conducted while participants engaged in a hippocampal-dependent[1,2] pattern separation task[1,2] (Fig. 1a). The task consisted of an encoding

and a retrieval phase. During encoding, subjects were shown 120 images, presented one at a time for 2-seconds. Each stimulus presentation was followed by a question prompt of 'indoor' or 'outdoor', whose duration was determined by subject response time via a button press. During retrieval, 160 images were presented serially and belonged to one of three categories: repeated images (40) from the encoding phase (repeat), similar but not identical images (80) to those presented during encoding (lure), and new images (40) (new). Following each 1-sec stimulus presentation period, a question prompt of 'new' or 'old' appeared, to which subjects responded with a button press. A 'new' response to a lure image (lure+) reflects successful discrimination between encoding and retrieval images. Conversely, an 'old' response to a lure image (lure-) reflects failed discrimination.

Participants

Electrophysiological recordings were acquired from eight adult (ages 23-52) patients (3 females) stereotactically implanted with depth electrodes (Integra or Ad-Tech, 5-mm interelectrode spacing) for clinical monitoring (Table 1). Surgical implantation was done to localize the seizure onset zone for possible surgical resection. Electrode placement was determined exclusively by clinical needs and took place at the University of California Irvine Medical Center. The inclusion criterion was concurrent electrode coverage over both hippocampal (HC) and neocortical (NC) areas. Informed consent was obtained from each participant for study inclusion in accordance with the institutional review board of the University of California, Irvine.

Electrophysiological recording acquisition and data pre-processing

Recording and offline filtering

The task was programmed in PsychoPy2 (Version 1.82.01). Participants performed the task on an Apple Macbook Pro placed at a comfortable viewing distance. Participants indicated

their responses using an external apple keypad. A photodiode was placed in the corner of the computer screen to capture stimulus presentation onset and offset timestamps with millisecond precision to time-lock them to the neural recordings.

Electrophysiological recordings were acquired using a Nihon Khoden recording system (256 channel amplifier, model JE120A), analog-filtered above 0.01 Hz and digitally sampled at 5000 Hz. Following acquisition, data were pre-processed using customized MATLAB scripts.

Post-acquisition, data were downsampled to 500 Hz. Line noise and harmonics were removed using a 2^{nd} order butterworth band stop infinite impulse response filter, with band stop frequencies centered around (+/-1Hz) 60Hz and 2^{nd}-3^{rd} harmonics.

Artefact rejection

Artefacts were identified in the time domain for each channel. Time points with values exceeding the mean channel signal +/- 3-4.5 multiples of the standard deviation were indicated. Timepoints near (1-sec before and after) the identified artefacts were also flagged as artefacts. After acquisition of spectral values, flagged artefactual timepoints were removed from the estimates of the mean and standard deviations of spectral power for baseline normalization. Moreover, any trial with a minimum of one artefact time point in the pre-stimulus baseline or post stimulus periods was removed entirely from all analyses.

Re-referencing

A Laplacian referencing scheme was implemented, whereby for a given electrode N, the averaged signal was computed over its neighboring electrodes, N-1 and N+1, and subtracted from the signal in electrode N. This has been previously suggested as an optimal scheme for slow frequency oscillatory activity[3]. However, all reported results were reproduced using two additional

approaches: 1) No post-acquisition re-referencing and 2) Using a probe-specific white-matter reference electrode.

Electrode localization and selection

Localization

Post implantation CT was co-registered to pre-implantation and post-implantation structural T1-weighted MRI scans for identification of electrode anatomical locations. There were two exceptions; for one patient, a pre-implantation T2-weighted MRI scan was used, and for a second patient, only a post-implantation T1-weighted MRI scan was used for identification.

Overlaid images were visualized in AFNI. An experienced rater (M. S. L.) identified electrodes in neocortical areas which included orbitofrontal (OFC), frontal (FRO), temporal (TEMP), cingulate (CING), insular (INS), and entorhinal/perirhinal (EC/PRC) cortical electrodes as well as hippocampal areas which included CA1, subiculum (SUB), and DG/CA3. OFC labeled electrodes are mutually exclusive from FRO electrodes. The first rater used a previously described[4] 0.55mm isotropic resolution in-house anatomical template, the Duvernoy atlas[5], and reported medial temporal lobe (MTL) localization guidelines[6] for anatomical identification of MTL electrodes. The labeled in-house template was resampled (1mm isotropic) and aligned to each subject's pre-implantation scans using ANTs Symmetric Normalization[7]. For all other non-MTL electrodes in the neocortex, Freesurfer cortical parcellations[8] and whole human brain atlas[9] were used.

Based on the anatomical labels within each subject's space, the electrode location was determined by identifying the region of interest that encompassed the center of the electrode artefacts. Cases in which electrodes were on the border between ROIs, bordering CSF, or between gray matter and white matter were noted as such and excluded from these analyses. A

notable exception from this exclusion criteria are the entorhinal/perirhinal localized electrodes that only bordered white matter or CSF, as electrodes localized in this region are few across all patients.

A second, experienced rater (M.A.Y.) independently localized entorhinal/perirhinal, hippocampal and DG/CA3. Across the two raters there was a 100% concordance for hippocampal sites, 100% concordance for DG/CA3 sites, and 75% concordance for entorhinal/perirhinal sites (those two sites were collapsed during analysis).

Electrode selection

The following inclusion criteria were applied for all electrodes (Fig. 1c): 1) Either located in the neocortex or the hippocampus, 2) placed entirely in gray matter (not overlapping with white matter or ventricles) 3) Cue responsive. Neocortical electrodes included: OFC, FRO, TEMP, CING, INS, EC/PRC. Hippocampal electrodes included: SUB, CA1, DG/CA3. Cue responsivity was defined as significant increases in slow frequency power (3-6 Hz) from pre-stimulus baseline (-0.3-0 sec vs. 0-1 sec) in any condition type (repeat+, lure-, lure+, new+), or across pooled trials from all condition types during the retrieval phase. Significant increases were measured using non-parametric, unpaired permutation testing (1000 permutations) between mean pre-stimulus and post-stimulus trial power values for each channel. Contacts from each subject meeting each of the inclusion criteria were projected onto a standard brain model (Montreal Neurological Institute, MNI) and visualized using DSI Studio. Retrieval cue responsive electrodes were used for all analyses (encoding and retrieval power and phase transfer entropy measures).

Spectral analysis

Power extraction and baseline normalization

Analytic Morlet wavelets were generated in MATLAB using the Wavelet toolbox and tested on simulated sinusoids of known spectral properties. Wavelets were used to extract power values in the 3-200 Hz frequency range, which were then converted to decibels (dB). For each trial, instantaneous power values were normalized relative to the channel mean and standard deviation of pre-stimulus power values at the same frequency. For a given trial, the pre-stimulus durations utilized to generate the normalizing distribution were only those preceding the same condition type as the trial to be normalized (condition specific pre-stimulus normalization). Then, for each trial, a z-score value was obtained for each time-frequency point; thereby generating a spectrogram. Baseline normalization was computed separately for the encoding and retrieval phases. Pre-stimulus artefact timepoints were included in the wavelet analysis to maintain the temporal structure of the data but were excluded from the normalizing distribution. A condition-specific baseline was utilized in this study due to condition-specific activity occurring in the response period, which precedes subsequent trials. However, results reported were reproduced with normalization relative to power distributions across the 1) entire recording, and 2) pre-stimulus baseline.

Group and individual subject analysis

For trial-type analysis, data were epoched relative to stimulus presentation and subject response timestamps. Z-score normalized spectrograms were averaged across trials within a channel to generate mean z-score normalized spectrograms for each channel. For region-based group spectral analysis, channels were pooled across subjects and defined as observations. For region-based individual subject analysis, spectrograms were averaged across channels in each trial, and trials were defined as observations.

Identification of significant time-frequency clusters

Significant time-frequency clusters which differed between the lure+ and lure- conditions were identified using cluster-based permutation testing[10]. Briefly, this involved calculating a t-statistic in each voxel, between the lure+ and lure- time-frequency z-score normalized matrices (z-maps), thereby generating the observed t-map. The observed t-map was then compared to a null distribution of t-maps generated over 1000 paired permutations. In each permutation, condition labels were shuffled within a random subset of channels (paired). A p-value for each voxel was obtained by comparing the observed to the null t-value in the same time-frequency voxel, thereby generating a p-map. To correct for multiple comparisons, clusters of contiguous voxels with a p <0.5 were identified and compared to the null-distribution cluster size. Observed clusters with sizes larger than the 95th percentile of those from the null distribution were considered significant after correction for multiple comparisons.

Phase transfer entropy analysis

An information theory metric, phase transfer entropy (PTE), was used to estimate the direction of information transfer in the 4-5 Hz frequency range.[11,12] PTE quantifies the mutual information between the past of signal 'X', X(t-τ) and the present of signal 'Y', Y(t). Mutual information reflects the degree of the reduction in uncertainty of one random variable (distribution of present phase values in region Y), given knowledge of another random variable (distribution of the past phase values in region X). PTE calculation was performed as described in Lobier et al., 2014[11] and normalized between -0.5 and 0.5 as described by Wang et al., 2019[12]. PTE was calculated using two methods; first as a time-average measure across the 2-second and 1-second periods following stimulus onsets in encoding and retrieval, respectively. Second, PTE was calculated in moving windows in these same periods (see PTE as a function of time (PTE FOT) below).

Briefly, PTE was calculated using the following formula,

$$PTE\ x \rightarrow y = H\big(Y(t) \mid Y(t-\tau)\big) - H(Y(t) \mid Y(t-\tau), X(t-\tau))$$

Where $H\big(Y(t) \mid Y(t-\tau)\big)$ is the conditional entropy of the distribution of $Y(t)$ conditioned on $Y(t-\tau)$, and $H(Y(t) \mid Y(t-\tau), X(t-\tau))$ is the conditional entropy of the distribution of $Y(t)$, conditioned on $Y(t-\tau)$ and $X(t-\tau)$).

The entropy of a random variable Y is,

$$H(Y) = -\sum_{y} p(y) \log_2 p(y)$$

and conditional entropy is calculated by,

$$H(Y|X) = -\sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{p(x)}$$

Given these definitions, the terms necessary to compute PTE $_{X\rightarrow Y}$ are the estimated marginal probability mass function (pmf) of the output past $Y(t-\tau)$, joint pmf of the output past and present $Y(t), Y(t-\tau)$, joint pmf of the output and input present $Y(t), X(t)$, and joint pmf of the output present, output past and input past $Y(t), Y(t-\tau), X(t-\tau)$. Each pmf was estimated using the entirety of the trial data, rather than a single trial. Bin width was defined according to Scott's choice[13] and the lag value τ was set to 100ms.

PTE $_{X\rightarrow Y}$ (NC → HC) and PTE $_{Y\rightarrow X}$ (HC → NC) were calculated separately and normalized as follows:

$$PTE = 0.5 - \left(\frac{PTE_{NC\rightarrow HC}}{PTE_{NC\rightarrow HC} + PTE_{HC\rightarrow NC}}\right)$$

Therefore, a value of 0.5 reflects equal bidirectional biases, and values larger and less than 0.5 reflect higher NC→HC and HC→ NC directional bias, respectively. Normalized PTE was then subtracted from 0.5, with a value of 0 denoting bidirectional interactions, a value greater and less than zero reflecting higher NC→HC and HC→ NC directional bias, respectively.

Phase transfer entropy as a function of time (PTE FOT)

PTE as a function of time was calculated as described above, however, in moving 0.5 sec windows with a 10ms step size (encoding and retrieval) and a 50ms step size (encoding).

Statistical analysis

After identification of the time-frequency cluster which significantly differed between lure+ and lure- across channels in a given region, power in this cluster was averaged in each channel, generating a single value. Mean channel cluster power values were used to statistically examine two additional contrasts, lure+ vs. new+ and lure+ vs. repeat+, using paired non-parametric permutations testing (randomly shuffling condition labels in a manner similar to what is described above), followed by false discovery rate (FDR) correction for multiple comparisons.

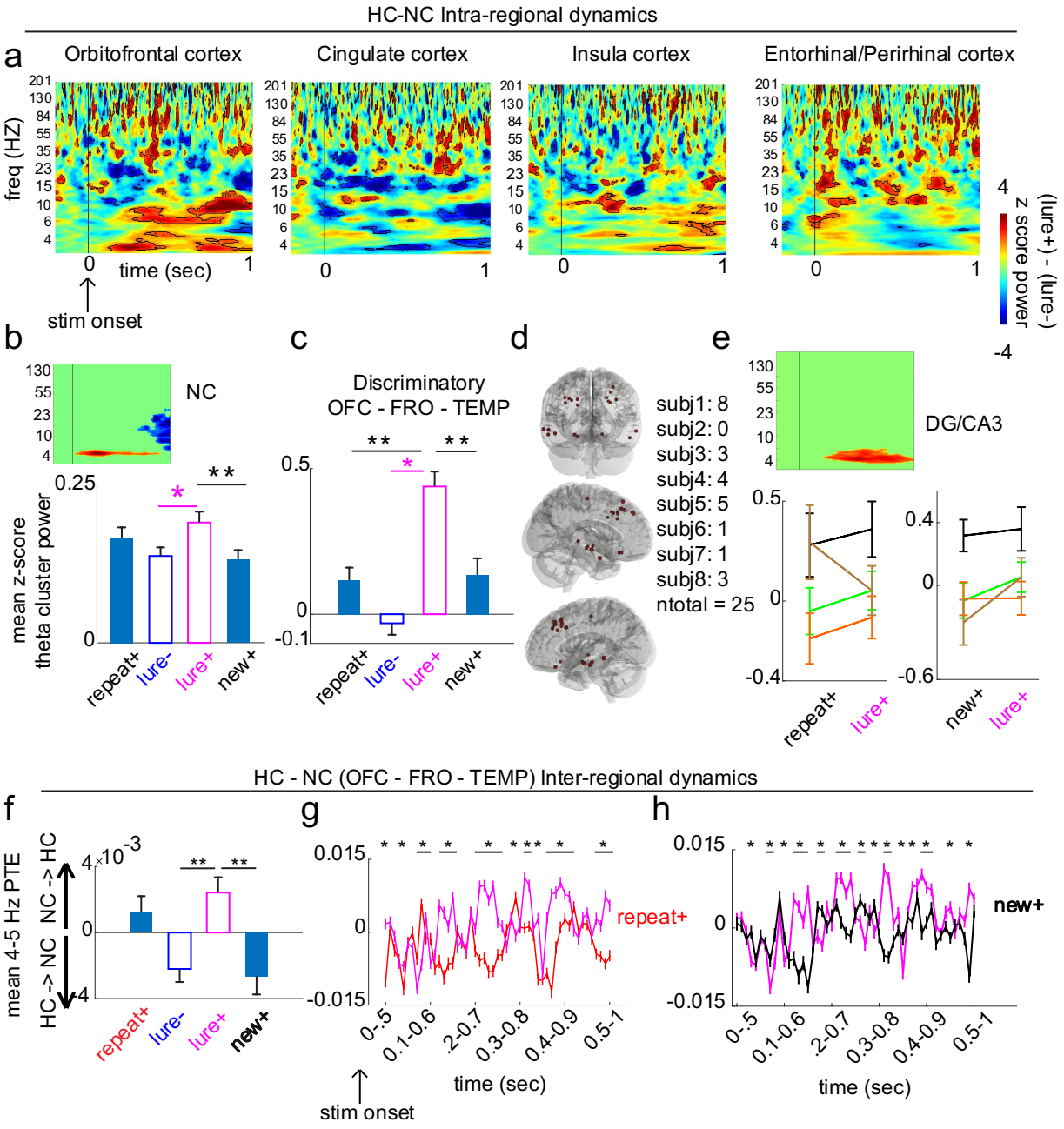
Similarly, paired non-parametric permutations testing was utilized for statistical examination of PTE values, using PTE values from channel NC-HC pairs as observations. A total of 1000 permutations were performed in all tests.

4.6 Supplementary material

Subject	OFC	FRO	TEMP	INS	CING	EC/PRC	DG/CA3	HC
1	2/4	9/14	5/7	0/1	2/2	1/2	2/3	8/11
2	0/5	2/5	0/4	0/0	2/4	2/5	0/0	1/5
3	2/4	15/18	7/12	6/13	3/6	0/1	2/4	4/9
4	5/7	6/7	9/11	0/0	2/6	3/7	2/3	8/13
5	2/2	13/13	8/11	0/0	7/9	0/0	1/3	7/13
6	3/4	8/11	5/11	1/3	1/1	2/3	02/1	2/6
7	2/2	4/4	8/17	2/2	2/4	1/2	0/0	6/6
8	1/1	10/10	7/13	4/6	2/3	0/0	0/0	4/4

Table 2. Proportion of task active electrodes per region during retrieval

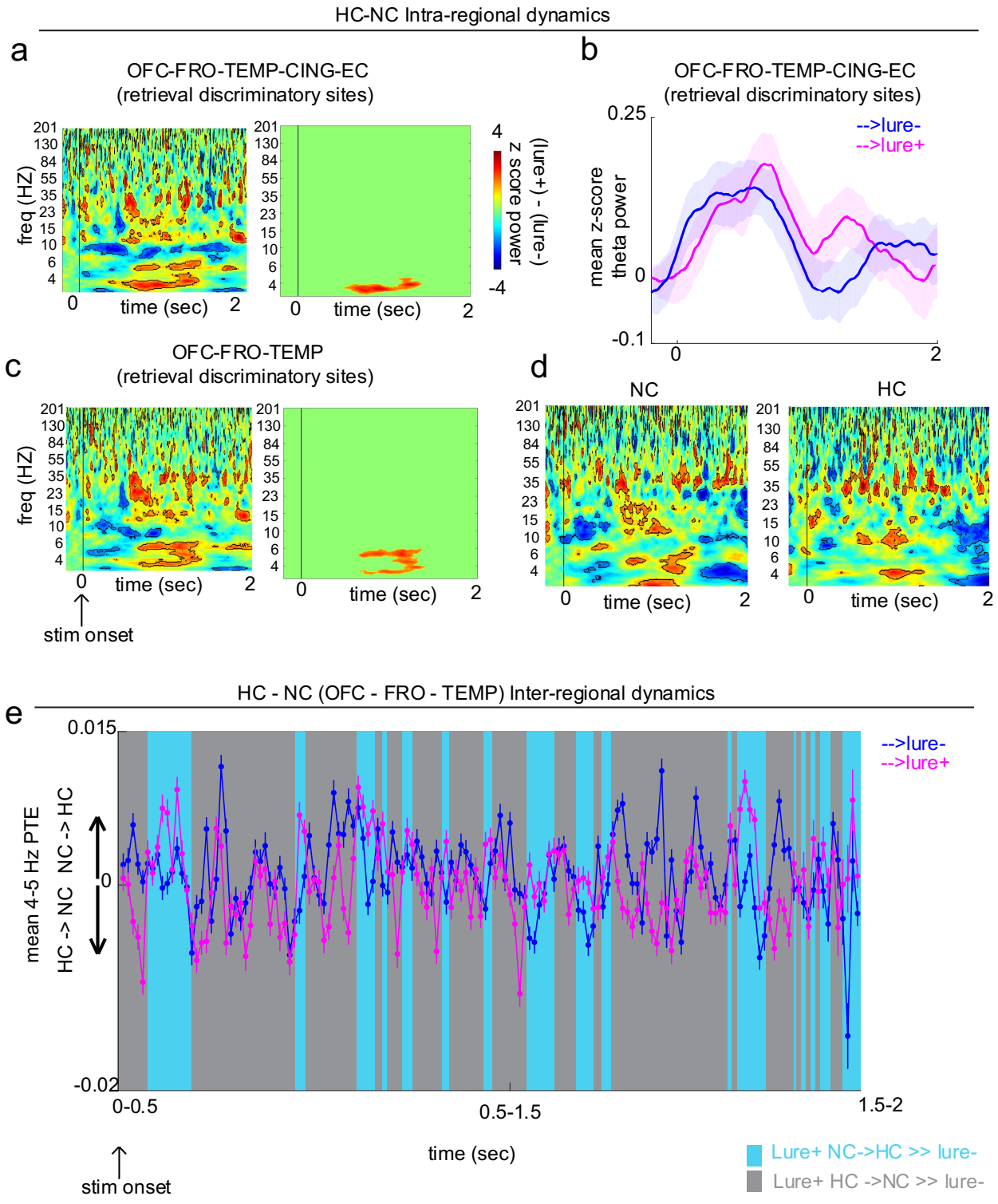
Cue responsivity was defined as a significant difference in slow frequency (3-6 Hz) power between pre and post stimulus presentation during retrieval. OFC: orbitofrontal cortex, FRO: frontal cortex, TEMP: temporal cortex, INS: insular cortex, CING: cingular cortex, EC/PRC: entorhinal/perirhinal cortex, DG/CA3: dentate gyrus / CA3; HC: hippocampus (includes DG/CA3, CA1 and subiculum).



Supplementary Figure S1. Hippocampal and neocortical intra-regional dynamics supporting during retrieval

a, Unthresholded z-score difference map (lure+ - lure-) in orbitofrontal (OFC), cingulate, insular, and entorhinal cortices maps (outlined contiguous voxels, $p < 0.05$ before multiple comparisons correction). Clusters were not significant after multiple comparisons correction. **b-c,** Mean NC (b) and discriminatory (lure+ vs. lure-) OFC, frontal (FRO), and temporal (TEMP) cortical sites (c) theta cluster power (inset) across four condition types. **b,** NC power was significantly higher in the lure+ compared to the new+ ($p = 0.002$) condition, but did not significantly differ from the repeat+ condition ($p = 0.1319$, $p_{(FDR)}$ threshold = 0.002) conditions. **c,** Discriminatory OFC, FRO and TEMP electrodes showed significantly higher theta power in the lure+ condition compared

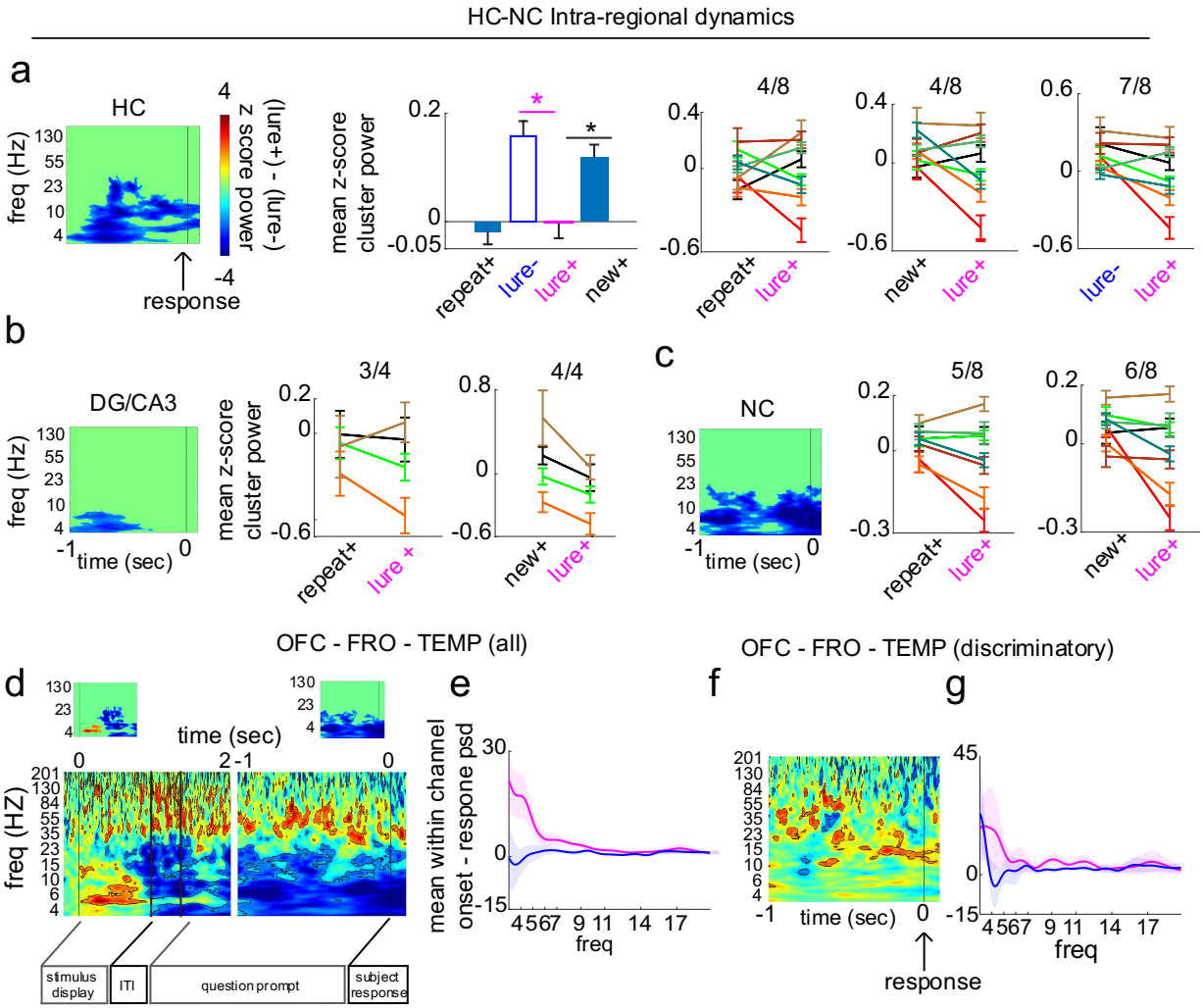
the repeat+ (p = 0.002) and the new+ conditions (p = 0.002, $p_{(FDR)}$ threshold = 0.002). **d,** Spatial map of OFC, FRO, and TEMP sites, in standardized (MNI) brain space, which showed significantly higher NC theta cluster power in the lure+ compared to the lure- condition (p <0.05). **e,** Individual subject mean theta cluster DG/CA3 power for repeat+ and new+ condition contrasts with lure+. **f-,** Phase transfer entropy (PTE) across the 1-sec post-stimulus period. A significantly higher weight to the NC → HC direction was observed in the lure+ compared to the lure- and new+ conditions ($p_{(FDR)}$ threshold = 0.002). **g-h** PTE as a function of time (FOT) (time x condition interaction p < 0.001 F (50,73400) = 19.631, and main effect for time F (50,73400) = 18.201, p < 0.001). PTE significantly differed during time epochs between the two conditions ($p_{(FDR)}$ threshold = 0.019). **h,** Same as in (g), but displaying lure+ and repeat+ contrasts (time x condition interaction F (50,73400) = 16.516, p < 0.001, p < 0.001, and, main effect for time F (50,73400) = 13.516, p < 0.001). PTE significantly differed in time epochs between the two conditions ($p_{(FDR)}$ threshold = 0.0220). Error bars represent the S.E.M across channels (power), channel pairs (PTE), or trials (individual subjects). *p < 0.05, **p < 0.01, ***p < 0.001 indicate significant differences using two-tailed non-parametric paired permutation testing defining channel/channel pairs as observations (1000 permutations).



Supplementary Figure S2. Hippocampal and neocortical intra- and inter-regional dynamics during encoding

a, Unthresholded (outlined contiguous voxels with $p < 0.05$, paired permutation testing, 1000 permutations) and thresholded (significant clusters after multiple comparisons correction, $p<0.05$)

z-score difference (→lure+ - →lure-) in all retrieval discriminatory NC sites. **b,** Mean theta (4-5 Hz) power traces across NC discriminatory sites. **c,** Same as in (a), but limiting retrieval discriminatory sites to OFC-FRO-TEMP. **d,** Unthresholded difference maps for all NC and all HC sites. **e,** PTE-FOT relative to stimulus onset, calculated over a 0.5 sec window and a 10ms overlap. Error shades and bars represent the S.E.M across channels (power) or channel pairs (PTE).


HC-NC Intra-regional dynamics
a
HC
freq (Hz)
130
55
23
10
4
response
4
-4
z score power (lure+) - (lure-)
mean z-score cluster power
0.2
0
-0.05
repeat+
lure-
lure+
new+
4/8
4/8
7/8
repeat+
lure+
new+
lure+
lure-
lure+
b
DG/CA3
time (sec)
3/4
4/4
repeat+
lure +
new+
lure+
c
NC
5/8
6/8
repeat+
lure+
new+
lure+
OFC - FRO - TEMP (all)
OFC - FRO - TEMP (discriminatory)
d
time (sec)
freq (HZ)
stimulus display
ITI
question prompt
subject response
e
mean within channel onset - respone psd
freq
f
g
time (sec)
response
HC - NC Inter-regional dynamics
h
3-20 Hz
i
4-5 Hz
NC - HC
0.001
0.001
probability
NC -> HC PTE
HC -> NC PTE
NC - HC
-0.004
-0.003
OFC-FRO-TEMP-CING-EC (discriminatory)
-0.003
0.005
OFC-FRO-TEMP (discriminatory)
-0.003
0.011

Supplementary Figure S3. Hippocampal and neocortical intra- and inter-regional dynamics during retrieval response period

a, HC slow frequency cluster (inset) power, relative to response, was significantly lower in the lure+ compared to the new+ condition ($p = 0.002$), but did not significantly differ from the repeat+ condition ($p = 0.6913$, $p_{(FDR)}$ threshold = 0.0279). Individual subject mean cluster power for repeat+ versus lure+, new+ vs. lure+, and lure- vs. lure+ contrasts are displayed. **b-c,** Individual subject mean slow frequency cluster power in DG/CA3 (c) and NC (d), relative to response, for repeat+ versus lure+, and new+ vs. lure+ contrasts. **d,** Z-score normalized unthresholded and thresholded (insets) difference maps (lure+ - lure-) locked to stimulus onset (left) and response (right) in NC OFC, FRO, and TEMP sites. **e, g,** Within-channel subtraction of the power spectral density (PSD) between the 1-sec post stimulus onset and 1-sec pre-response period (onset - response PSD) in the indicated sites. **f,** z-score unthresholded difference map locked to response in OFC, FRO, and TEMP discriminatory sites. **h,i,** Probability density function of within channel onset minus response (onset - response) NC–> HC (cyan) and HC→ NC (black) PTE. Distributions are displayed for PTE computed across 3-20Hz in steps of 1-Hz (h), in the 4-5Hz range (i) in the indicated areas. Error shades and bars represent the S.E.M across channels (power) or trials (individual subjects). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ indicate significant differences using two-tailed non-parametric paired permutation testing defining channel/channel pairs as observations (1000 permutations).

4.7 References

1. Greenstein YJ, Pavlides C, Winson J. Long-term potentiation in the dentate gyrus is preferentially induced at theta rhythm periodicity. *Brain Research*. 1988;438(1-2):331-334.

2. Larson J, Lynch G. Induction of synaptic potentiation in hippocampus by patterned stimulation involves two events. *Science*. 1986;232(4753):985-988.

3. Nyhus E, Curran T. Functional role of gamma and theta oscillations in episodic memory. *Neuroscience and Biobehavioral Reviews*. 2010;34(7):1023-1035.

4. Lisman JE, Idiart MAP. Storage of 7 ± 2 short-term memories in oscillatory subcycles. *Science*. 1995;267(5203):1512-1515.

5. Hyman JM, Wyble BP, Goyal V, Rossi CA, Hasselmo ME. Stimulation in Hippocampal Region CA1 in Behaving Rats Yields Long-Term Potentiation when Delivered to the Peak of Theta and Long-Term Depression when Delivered to the Trough. *Journal of Neuroscience*. 2003;23(37):11725-11731.

6. Teyler T.J., DiScenna P. The hippocampal memory indexing theory. *Behavioral neuroscience*. 1986;100(2):147-154.

7. Mcclelland JL. Incorporating Rapid Neocortical Learning of New Schema-Consistent Information into Complementary Learning Systems Theory. 2013;142(4):1190-1210.

8. Karimi Abadchi J, Nazari-Ahangarkolaee M, Gattas S, et al. Spatiotemporal patterns of neocortical activity around hippocampal sharp-wave ripples. *eLife*. 2020;9.

9. Yassa MA, Stark CEL. Pattern separation in the hippocampus. *Trends in Neurosciences*. 2011;34(10):515-525.

10. Buzsáki G. Theta rhythm of navigation: Link between path integration and landmark navigation, episodic and semantic memory. *Hippocampus*. 2005;15(7):827-840.

11. Hasselmo ME, Stern CE. Theta rhythm and the encoding and retrieval of space and time. *NeuroImage*. 2014;85:656-666.

12. Zheng J, Stevenson RF, Mander BA, et al. Multiplexing of Theta and Alpha Rhythms in the Amygdala-Hippocampal Circuit Supports Pattern Separation of Emotional Information Article Multiplexing of Theta and Alpha Rhythms in the Amygdala-Hippocampal Circuit Supports Pattern Separation of Emotional Information. *Neuron*. 2019:1-12.

13. Hsieh LT, Ranganath C. Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval. *NeuroImage*. 2014;85:721-729.

14. Fries P. A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*. 2005;9(10):474-480.

15. Guderian S, Düzel E. Induced theta oscillations mediate large-scale synchrony with mediotemporal areas during recollection in humans. *Hippocampus*. 2005;15(7):901-912.

16. Griffiths BJ, Parish G, Roux F, et al. Directional coupling of slow and fast hippocampal gamma with neocortical alpha/beta oscillations in human episodic memory. *Proceedings of the National Academy of Sciences of the United States of America*. 2019;116(43):21834-21842.

17. Teyler TJ, Rudy JW. The hippocampal indexing theory and episodic memory: Updating the index. *Hippocampus*. 2007;17(12):1158-1169.

18. McClelland O, McNaughton BL, O'Reilly RC. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*. 1995;102(3):419-457.

19. McClelland JL, Mcnaughton BL, Lampinen AK. Integration of new information in memory: new insights from a complementary learning systems perspective *Philos Trans R Soc Lond B Biol Sci*. 2020;375(1799):20190637.

20. Norman KA, O'Reilly RC. Modeling Hippocampal and Neocortical Contributions to Recognition Memory: A Complementary-Learning-Systems Approach. *Psychological Review*. 2003;110(4):611-646.

21. Norman KA. Revisiting the Complementary Learning Systems model. *Hippocampus*. 2010;20(11):1217-1227.

22. Colgin LL. Rhythms of the hippocampal network. *Nature Reviews Neuroscience*. 2016;17(4):239-249.

23. Hsieh LT, Ekstrom AD, Ranganath C. Neural oscillations associated with item and temporal order maintenance in working memory. *Journal of Neuroscience*. 2011;31(30):10803-10810.

24. Roberts BM, Hsieh LT, Ranganath C. Oscillatory activity during maintenance of spatial and temporal information in working memory. *Neuropsychologia*. 2013;51(2):349-357.

25. Colgin LL, Denninger T, Fyhn M, et al. Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature*. 2009;462(7271):353-357.

26. Manns JR, Zilli EA, Ong KC, Hasselmo ME, Eichenbaum H. Hippocampal CA1 spiking during encoding and retrieval: Relation to theta phase. *Neurobiology of Learning and Memory*. 2007;87(1):9-20.

27. O'Keefe J, Recce ML. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*. 1993;3(3):317-330.

28. Hafting T, Fyhn M, Bonnevie T, Moser MB, Moser EI. Hippocampus-independent phase precession in entorhinal grid cells. *Nature*. 2008;453(7199):1248-1252.

29. Burgess N, O'Keefe J. Models of Place and Grid Cell Firing and Theta Rhythmicity. Europe PMC Funders Group. *Curr Opin Neurobiol*. 2011;21(5):734-744.

30. Lenck-Santini PP, Fenton AA, Muller RU. Discharge properties of hippocampal neurons during performance of a jump avoidance task. *Journal of Neuroscience*. 2008;28(27):6773-6786.

31. McClelland JL. Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General*. 2013;142(4):1190-1210.

32. Tse D, Takeuchi T, Kakeyama M, et al. Schema-dependent gene activation and memory encoding in neocortex. *Science*. 2011;333(6044):891-895.

33. Tse D, Langston RF, Kakeyama M, et al. Schemas and memory consolidation. *Science*. 2007;316(5821):76-82.

34. Migo E, Montaldi D, Norman KA, Quamme J, Mayes A. The contribution of familiarity to recognition memory is a function of test format when using similar foils. *Quarterly Journal of Experimental Psychology*. 2009;62(6):1198-1215.

35. Lohnas LJ, Duncan K, Doyle WK, Thesen T, Devinsky O, Davachi L. Time-resolved neural reinstatement and pattern separation during memory decisions in human hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*. 2018;115(31):E7418-E7427.

36. Jutras MJ, Fries P, Buffalo EA. Oscillatory activity in the monkey hippocampus during visual exploration and memory formation. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(32):13144-13149.

37. Kaplan R, Doeller CF, Barnes GR, et al. Movement-related theta rhythm in humans: Coordinating self-directed hippocampal learning. *PLoS Biology*. 2012;10(2).

38. Yassa MA. Brain Rhythms: Higher-Frequency Theta Oscillations Make Sense in Moving Humans. *Current Biology*. 2018;28(2):R70-R72.

39. Aghajan ZM, Schuette P, Fields TA, et al. Theta Oscillations in the Human Medial Temporal Lobe during Real-World Ambulatory Movement. *Current Biology*. 2017;27(24):3743-3751.e3.

40. Addante RJ, Watrous AJ, Yonelinas AP, Ekstrom AD, Ranganath C. Prestimulus theta activity predicts correct source memory retrieval. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108(26):10702-10707.

41. Guderian S, Schott BH, Richardson-Klavehn A, Düzel E. Medial temporal theta state before an event predicts episodic encoding success in humans. *Proceedings of the National Academy of Sciences of the United States of America*. 2009;106(13):5365-5370.

42. Norman KA, Newman E, Detre G, Polyn S. How inhibitory oscillations can train neural networks and punish competitors. *Neural Computation*. 2006;18(7):1577-1610.

43. Colgin LL. Mechanisms and Functions of Theta Rhythms. *Annual Review of Neuroscience*. 2013;36(1):295-312.

44. Amilhon B, Huh CYL, Manseau F, et al. Parvalbumin Interneurons of Hippocampus Tune Population Activity at Theta Frequency. *Neuron*. 2015;86(5):1277-1289.

45. Bonifazi P, Goldin M, Picardo MA, et al. GABAergic hub neurons orchestrate synchrony in developing hippocampal networks. *Science*. 2009;326(5958):1419-1424.

46. Titiz AS, Hill MRH, Mankin EA, et al. Theta-burst microstimulation in the human entorhinal area improves memory specificity. *eLife*. 2017;6:1-18.

47. Suthana N, Haneef Z, Stern J, et al. Memory enhancement and deep-brain stimulation of the entorhinal area. *New England Journal of Medicine*. 2012;366(6):502-510.

48. Eliav T, Geva-Sagiv M, Yartsev MM, et al. Nonoscillatory Phase Coding and Synchronization in the Bat Hippocampal Formation. *Cell*. 2018;175(4):1119-1130.e15.

49. Bush D, Burgess N. Advantages and detection of phase coding in the absence of rhythmicity. *Hippocampus*. 2020; doi: 10.1002/hipo.23199.

Chapter 5: Two methods for estimating directional communication between brain areas (deductive)

(1) Time-dependent phase transfer entropy and (2) General Linear Kalman Filter based estimates for Granger causality

5.1 Original definition of phase transfer entropy

In the last chapter, a metric called phase transfer entropy (PTE, Lobeir et al, 2014) was used to make inferences about the direction of communication between two memory systems in the brain – the hippocampus and the neocortex. In this chapter, I emphasize how this metric was adapted to extend it to a time dependent measure.

First, a brief description is provided regarding the original definition and use of PTE. For a given pair of simultaneously recorded time series (ex. two electrode sensors recording from two different brain areas), the data is first band-pass filtered for a desired frequency range. Then, the phase signal of such frequency range is obtained. PTE quantifies the mutual information between the past of phase signal 'X', X(t-τ) from one electrode sensor and the present of phase signal 'Y', Y(t) from another electrode sensor.

$$PTE\ x \rightarrow y = H\big(Y(t) \,\big|\, Y(t-\tau)\big) - H(Y(t) \mid Y(t-\tau), X(t-\tau))$$

Where $H\big(Y(t) \,\big|\, Y(t-\tau)\big)$ is the conditional entropy of the distribution of $Y(t)$ conditioned on $Y(t-\tau)$, and $H(Y(t) \mid Y(t-\tau), X(t-\tau))$ is the conditional entropy of the distribution of $Y(t)$, conditioned on $Y(t-\tau)$ and $X(t-\tau)$).

The entropy of a random variable Y is,

$$H(Y) = -\sum_{y} p(y) \log_2 p(y)$$

and conditional entropy is calculated by,

$$H(Y|X) = -\sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{p(x)}$$

Given these definitions, the terms necessary to compute PTE $_{X\rightarrow Y}$ are the estimated marginal probability mass function (pmf) of the output past $Y(t-\tau)$, joint pmf of the output past and present $Y(t), Y(t-\tau)$, joint pmf of the output and input present $Y(t), X(t)$, and joint pmf of the output present, output past and input past $Y(t), Y(t-\tau), X(t-\tau)$.

In the context of multi-trial experimental design where several observations are collected to probe a given cognitive function, the entirety of the trial data is used to generate the aforementioned pmfs. For example, PTE can be calculated using n 1-sec trials. Depending on the sampling rate (fs), the data used to generate the pmf would be nTrials*tTimepoitns*fs values. Such data indexing would provide a single PTE measure for the 1-sec period, using timepoints from all trials, and defining the timepoint as the observation.

An example from chapter 4 is provided, whereby a single PTE is calculated using 1-sec trials, separately for two conditions (correct and incorrect subject response; lure- and lure+, respectively, **Fig. 1**).

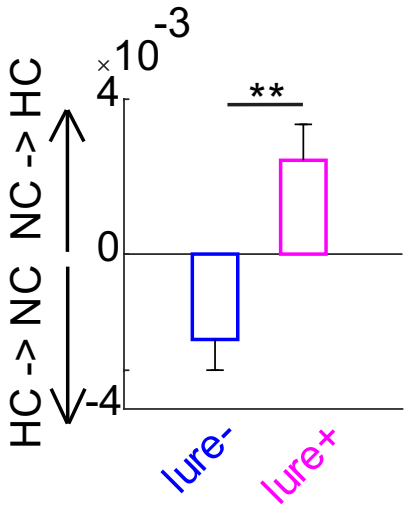


Figure 1. Example application of PTE calculated over a single time epoch

(see chapter 4 for more detailed figure information)

5.2 Extension of the PTE metric into a time-dependent one

The PTE measure can be adapted to a time dependent one. PTE can be calculated in moving windows of shorter duration (ex. 0.5 sec) and with a small step size (ex. 10ms or 50 ms). The pmf generated for a time-dependent metric would be estimated separately for each time window, using datapoints only from the time window of interest (ex – the first window would be 0-0.5 sec), then iteratively repeating the PTE calculation for each moving window. This would then yield a time dependent PTE metric (**Fig. 2**).

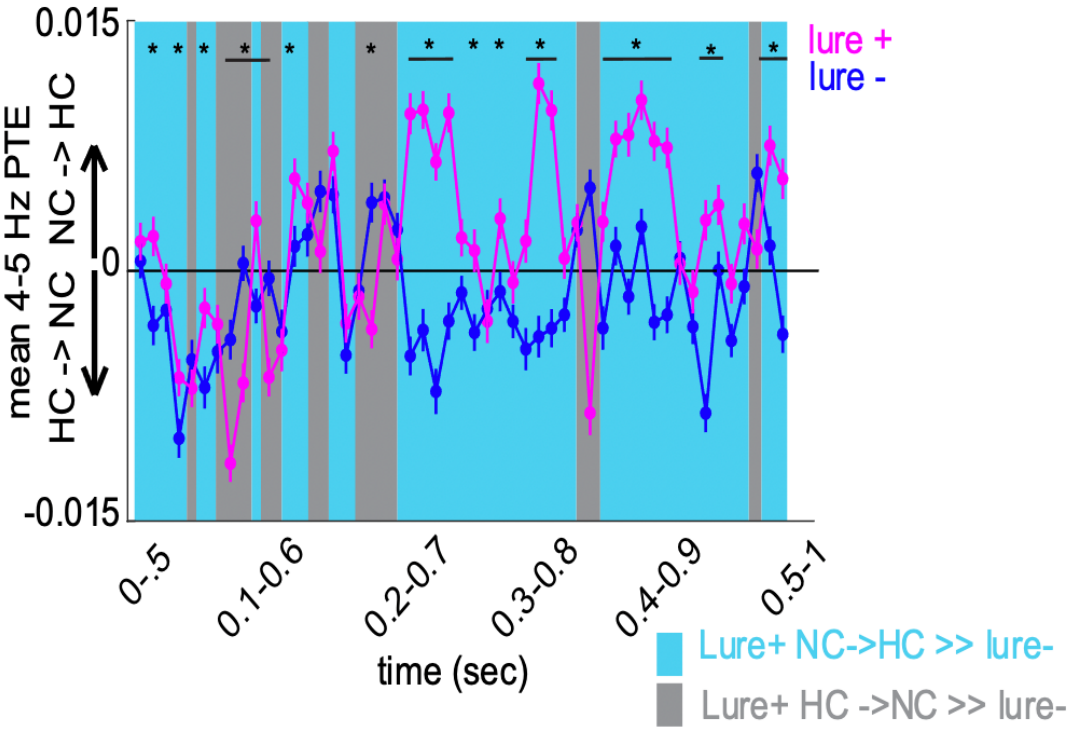


Figure 2. Example application of time dependent PTE

(see chapter 4 for more detailed figure information)

5.3 Considerations regarding time-dependent PTE estimates

A crucial step to the PTE calculation is the probability mass function estimates. With smaller time windows, less observations (time-points) are available for such estimates. Therefore, the choice of a time-window (and utilized fs) should account for the trade-off between temporal granularity and adequate pmf estimates. Qualitative approaches for evaluating the adequacy of the pmf estimates include comparing the variability in PTE output measures for small step sizes between windows (large % overlap). Additionally, the time dependent PTE measure

(mean PTE across all time windows) should be consistent with the single PTE calculated over the entire period. Both qualitative tests are met in the examples provided above (**Figs. 1-2**)

5.4 Advantage of time-dependent PTE in the context of neural data analysis

The brain operates at a wide range of time scales, exhibiting ultraslow (0.01 to 0.1) to ultrafast (400-800 Hz) frequencies (Hughes 2008, Watson 2018). Behavioral neuroscience experiments occurring while neural data is simultaneously collected traditionally have a trial/condition based design, whereby each trial is of particular duration relevant for the behavior. Since behavioral executions occur at slower time scales compared to underlying neural processes that support such behavior, having a fine-grained (faster time scales) analysis of neural processes is critical.

In the examples provided above, the single PTE metric calculated using 1-sec data for the lure+ condition (successful memory performance) suggests there is larger PTE for the neocortex → hippocampus direction (**Fig 1**). However, the temporally fine-grained analysis using time-dependent PTE shows that the directional communication is largely *bi*directional; sometimes the hippocampus → neocortex direction is larger than the neocortex → hippocampus, and vice versa (**Fig 2**). This finding of bidirectionally of communication between these two memory systems was meaningful given a prior unidirectionality hypothesis in the field. Lastly, this approach informs as to where in the 1-sec trial directional biases occur; in the current example, the neocortex → hippocampal directional bias predominated specifically in the later period of the trial. This guided the interpretation of the single PTE metric calculated in Figure 1, regarding what kind of information (ex. perceptual vs. mnemonic) may be communicated to the hippocampus by the neocortex.

5.5 References

Hughes JR. Gamma, fast, and ultrafast waves of the brain: their relationships with epilepsy and behavior. Epilepsy Behav. 2008 Jul;13(1):25-31. doi: 10.1016/j.yebeh.2008.01.011. Epub 2008 Apr 24. PMID: 18439878.

Lobier M, Siebenhühner F, Palva S, Palva JM. Phase transfer entropy: a novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions. Neuroimage. 2014 Jan 15;85 Pt 2:853-72. doi: 10.1016/j.neuroimage.2013.08.056. Epub 2013 Sep 2. PMID: 24007803.

Watson BO. Cognitive and Physiologic Impacts of the Infraslow Oscillation. Front Syst Neurosci. 2018 Oct 16;12:44. doi: 10.3389/fnsys.2018.00044. PMID: 30386218; PMCID: PMC6198276.

5.6 Kalman filter-based granger causality for estimating directional communication between the hippocampus and the neocortex. Abstract

We aim to examine the directionality of communication between hippocampus and neocortex, using the same data as in Chapter 4, using an alternate approach. Here, we used the General Linear Kalman filter to estimate the coefficients of a multi-variable, time-varying autoregressive model (MV-TV- AR). Subsequently, we implemented time-varying partial granger causality (tv-pGC), which incorporates the measurement residual and measurement prediction covariance, to infer temporal relationships between brain areas. We found that the estimated model can successfully predict a given measurement point, using a linear combination of past measurement values weighted by the estimated AR coefficients. Moreover, we found an increase in GC index, reflecting directional communication from the hippocampus to the neocortex, preceding successful discrimination between highly similar memories.

5.7 Kalman filter-based granger causality for estimating directional communication between the hippocampus and the neocortex. Introduction

We aim to model the human intracranial multi-trial, multichannel data during a pattern separation task (see Chapter 3 for a detailed description of the experimental design and associated data) using a time-varying multivariate autoregressive (MV-TV- AR) model. This

model will then be used in a Granger causality calculation to enable for the investigation of directional linear dependence between neocortex and hippocampus that support pattern separation [1].

Given that the spectral content of the data as well as the interaction of neural ensembles change with time, and the large number of implanted electrodes, we will utilize a time-varying multivariate model, respectively [2]. The model is estimated using a General Linear Kalman Filter (GLKF), which is similar to the classical Kalman filter counterpart, however, it differs in that it extends the state space representation to a matrix state space model for a multi-trial, multivariate time series [3]. In comparison to other estimators, Kalman filtering has gained acceptance for its accurate instantaneous estimation of time-varying AR parameters modeling high dimensional data [2]. In regard to data with high signal-to-noise ratio, and large number of electrodes and trial numbers, it was shown that the GLKF outperforms the classical one [2]. Subsequent to estimation of the TV-MVAR model parameters by means of the GLKF, a measure of linear dependence between brain regions, the granger causality index (GCI) is calculated.

Using the AR modeling by means of GLKF followed by GCI, we observed an effective functional connectivity in the hippocampal- neocortical network as well as within the hippocampal network that may support successful discrimination between highly similar memories.

5.8 Kalman filter-based granger causality for estimating directional communication between the hippocampus and the neocortex. Methods

Time-varying vector AR Model Fitting

A time-varying, vector autoregressive (VAR) model is defined by

$$Y(n) = \sum_{k=1}^{p} A_k(n)\, Y(n-k) + e(n) \tag{1}$$

Where $Y(n)_{(\text{nelec})}$is a vector,

$A_{k_{(nelec,nelec)}}$ is a time-varying matrix, and

$nelec$ represents the number of electrodes, n represents the time index, and p

represents the model order.

The model order was chosen such that the root mean squared error (RMSE) is reduced, while avoiding large p for computational ease and prevention of model overfitting (**Fig. 1**).

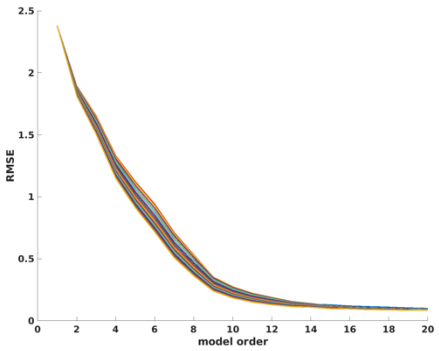


Fig. 1. Choice of model order
Average RMSE across all trials and all electrodes between the predicted measurement by the estimated AR model coefficients and the true measurement. Each line represents a different value of parameter c in equation 6.

The General Linear Kalman Filter

Implementation of the GLKF filter was based on the reported Milde, et al. approach [3].

The GLKF operates based on the state (dynamic) and measurement equations defined in

equations 2 and 3, respectively. The state equation is defined by

$$X(k+1) = F(k)X(k) + v(k) \quad k = 1,2,..,N \tag{2}$$

$F(k)_{(\text{p,nelec})}$ is the state transition matrix defined as an identity (I) matrix.
$X(k)_{(\text{p*nelec,nelec})}$ is the state being estimated which is *p* AR model coefficients. $X(k)$ is a time-varying matrix.
$v(k)$ is the process noise, whose covariance was defined as $CI_{(\text{p,nelec})}$ where C was set to 1.

The measurement equation (3) is defined based on equation 1.

$$Z(k) = H(k)X(k) + w(k) \tag{3}$$

$$H(k) = [D(k-1)\, D(k-2) \ldots D(k-p)] \tag{4}$$

$H(k)_{(\mathrm{ntrials,p*nelec})}$ contains concatenated neural data from matrix D_{k} which is a 2-D matrix for timepoint *k* from $D_{(\mathrm{ntrials,nelecs,ktimepoints})}$

$$(D_k)_{k=1}^{N} \begin{bmatrix} D_1^1(k) & \cdots & D_{nelec}^1(k) \\ \vdots & \ddots & \vdots \\ D_1^{ntrials}(k) & \cdots & D_{nelec}^{ntrials}(k) \end{bmatrix}_{k=1}^{N} \tag{5}$$

The recorded data is modeled such that for a given electrode, the trials are a collection of realizations from a stochastic process. Moreover, all electrodes reflect an underlying multivariate stochastic process. $H(k)_{(\mathrm{ntrials,p*nelec})}$ is *p* past time points from $D_{(\mathrm{ntrials,nelecs,ktimepoints})}$, whose multiplication by the estimated coefficients $X(k)_{(\mathrm{p*nelec,nelec})}$ generate the predicted $Z(k)$.

$w(k)$is the measurement noise with covariance matrix $R(k)$ which is defined below [4,5].

$$R(k) = R(k-1)(1-c) + \frac{c\big(Z(k) - Z(k|k-1)\big)'(Z(k) - X(k|k-1))}{ntrials - 1} \tag{6}$$

The constants C and c for the estimation of process and measurement noise, respectively, were set to 1 for two reasons. First, the RMSE between the recorded neural signal and that predicted by the estimated AR model decreased as a function of increased c (**Fig. 2**). Moreover, setting a higher value for c increases the weight in the $R(k)$ equation for the current measurement prediction errors rather than previous errors. This is reasonable given that we are trying to estimate the error in the measurement rather than that of the filter. Since filter performance improves with time, adding a larger weight to the current error rather than a

weighted combination of recent and past errors, will more likely capture error in the measurement.

The following initial conditions were defined:

$X(p) = 0, R(p) = I_{nelec}, P(p) = I_{(nelec,p)}$

Following the above definitions, the classical KF algorithm was iteratively repeated from $k = P + 1 : N$ as follows:

Predicted state

$$X(k+1|k) = F(k)X(k|k) \tag{8}$$

State prediction covariance

$$P(k+1|k) = F(k)P(k)F(k)' + Q(k) \tag{9}$$

Predicted measurement

$$Z(k+1|k) = H(k+1)X(k+1|k) \tag{10}$$

Measurement prediction covariance

$$S(k+1) = H(k+1)P(k+1|k)H(k+1)' + trace\big(R(k+1)\big)I_{ntrials} \tag{11}$$

Filter gain

$$W(k+1) = P(k+1|K)H(k+1)'S^{-1}(k+1) \tag{12}$$

Updated state estimate

$$X(k+1|k+1) = X(k+1|k) + W(k+1)v(k+1) \tag{13}$$

Where $v(k+1) = Z(K+1) - Z(k+1|k)$ is the measurement residual, (14)

and $Z(k+1) = D(K+1)$ is the recorded neural data. (15)

Joseph form covariance update

$$P(k+1|k) = \big(I - W(k+1)H(k+1)\big)\, p(k+1|k)\big(I - W(k+1)H(k+1)\big)' + \big(W(k+1)\, trace\big(R(k+1)\big)I_{ntrials}W(k+1)'\big)$$

Filter performance was evaluated based on RMSE (**Fig. 2, 3**) of the measured signals versus the signal predicted by the estimated AR model.

Time Varying Partial Granger Causality Index

Granger causality index (GCI) is a measure of the degree to which information in a given signal can be explained by the past of another signal. In order to calculate time varying-partial GCI (tv-pGCI), a ratio of the prediction error term from the estimation of AR model is used. Since the AR model cannot perfectly explain the time series, at each time iteration k of the Kalman filter, there is an associated prediction error

$PE(k) = R(k)$, where

$$R(k) = R(k-1)(1-c) + \frac{c\left(Z(k)-Z(k|k-1)\right)\left(Z(k)-Z(k|k-1)\right)'}{ntrials-1}$$

First, $PE(k)$ is calculated with all electrodes included. Then, $PE_{reduced}(k)$ is calculated each time with a single electrode removed.

$$tv, pGCI_{\ i\rightarrow j} = ln\left(\frac{PE_{reduced}(i,j)}{PE(j,j)}\right) \quad (17)$$

Focusing on the diagonal elements, a ratio of $PE_{reduced}(j, j)$ (with electrode i removed) to PE(j,j) (with all electrodes included) indicates whether there was a reduction or an increase in prediction error of the signal in electrode j, with electrode i removed. A tv-pGCI ratio greater than zero, indicates a directional relationship from signal i to signal j, in that prediction error of signal j is reduced with the inclusion of electode i. GCI is calculated between all possible pairs of electrodes, and as a function of time.

Intracranial recordings and memory task data acquisition

The data utilized for this study is from a human subject who underwent a neurological surgery for implantation of electrodes in the brain for clinical monitoring. Local field potential recordings were acquired from over 100 electrodes covering brain regions including neocortex and hippocampus. Recordings were acquired at a sampling rate of 5000 Hz while patients engaged in a task that required discrimination between highly similar memories.

5.9 Kalman filter-based granger causality for estimating directional communication between the hippocampus and the neocortex. Results

General Linear Kalman Filter and AR Model Performance

Proper GLKF performance would indicate proper estimation of the AR model coefficients. We tested the accuracy of the estimation of the AR model coefficients by applying equation 1, in that we used the model to predict a current value of the signal given a linear combination of past values of the signal weighted by the coefficients. Figure 2 shows an exemplar trace of the true signal compared to that of the predicted signal by the model. The average RMSE across trials and electrodes between the measured and predicted signals is 0.17 – a value that is over 257 times smaller than the RMS of the data (43.396) (**Fig. 4**).

We used the same procedure of average RMSE calculation to define the model order and filter parameter c (**Fig. 2, 3**). We found that average RMSE decreases with model order for any parameter c (**Fig 1**). Therefore, the model order was set to 10 given the trade-off between RMSE and computation time. Subsequently, different values of c were examined for the model order 10. This parameter is part of equation 6 for $R(k)$ and $Q(k)$ estimation – an approach previously applied to neural data [3]. The parameter c was set to 1 for a decreased RMSE and as well as to add a higher weight to current error rather than previous errors in the GLKF algorithm.

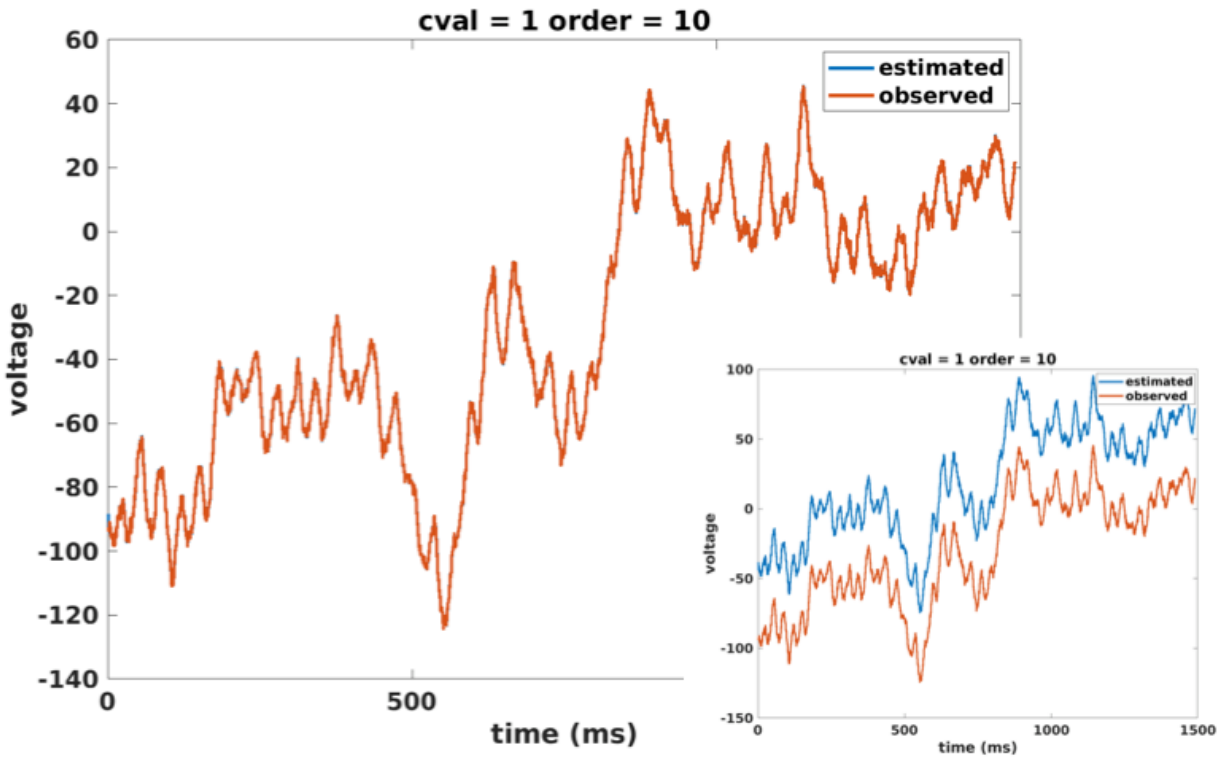


Figure. 2. Examination of filter and model performance
The main figure shows an overlay of the observed signal (recorded neural signal) in orange and the predicted signal by the estimated AR model coefficients in blue. Since there is high correspondence between the estimated and observed, the two signals are plotted in the inset with a +50 y-axis shift for better visualization.

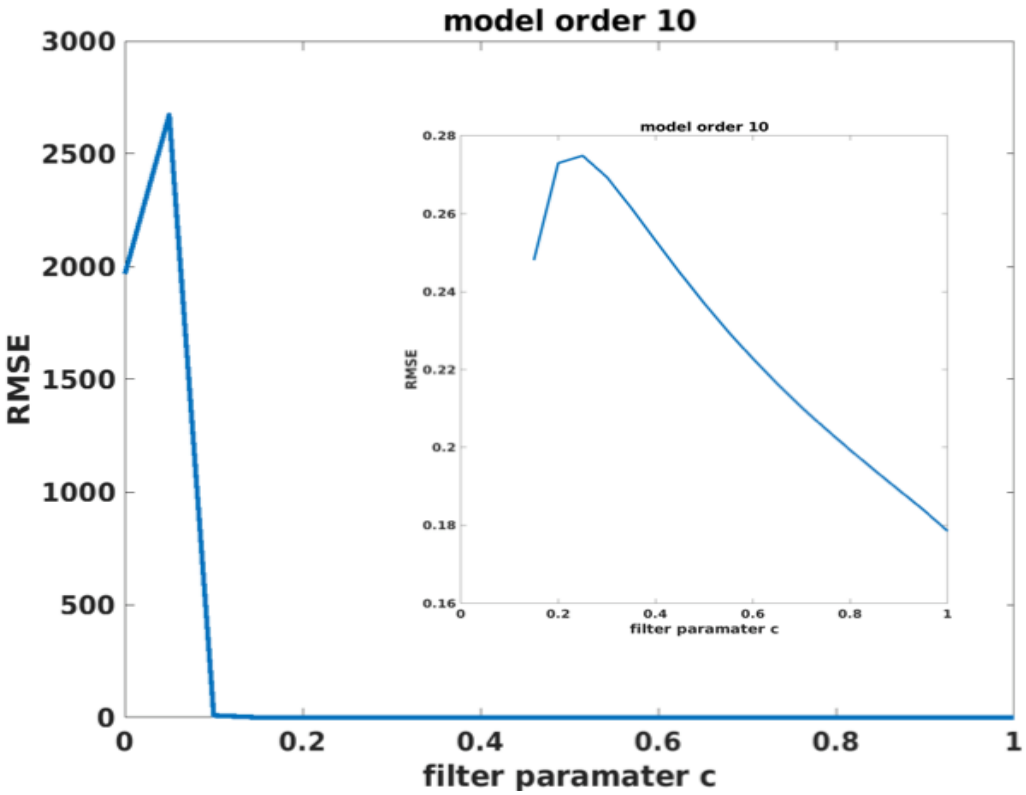


Figure 3. Definition of filter parameter
Average RMSE across all trials and electrodes between the observed measurement and that predicted by the estimated AR model coefficients as a function of filter parameter c. Significantly large errors are observed for values of c <0.15. Shown in the inset figure, is the average RMSE for values >0.15.

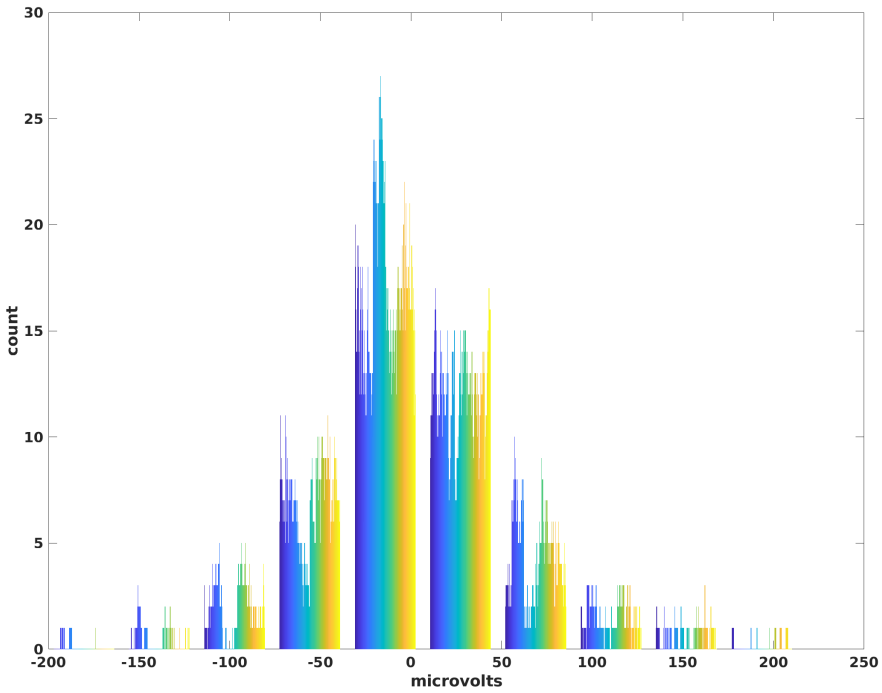


Figure 4. Distribution of voltage values of recorded data

Tv-pGCI

In order to investigate the directionality of interaction between different brain areas during successful discrimination between highly similar memories, tv-pGCI was computed. More specifically, the direction of interaction between different neocortical sites and the hippocampus was examined during two behavioral conditions: successful discrimination (pattern separation) and failure to discriminate (pattern completion) between similar memories. Using data from one patient, observed is a decrease in GCI from hippocampal to neocortical electrodes and between different hippocampal electrodes in the 0.5 sec prior to display of a highly similar memory that is later incorrectly discriminated (**Fig. 5a-b**). However, this decrease in GCI is not observed preceding the trials where successful discrimination occurs. This effect was observed between cingulate cortex and prefrontal cortex with the hippocampus.

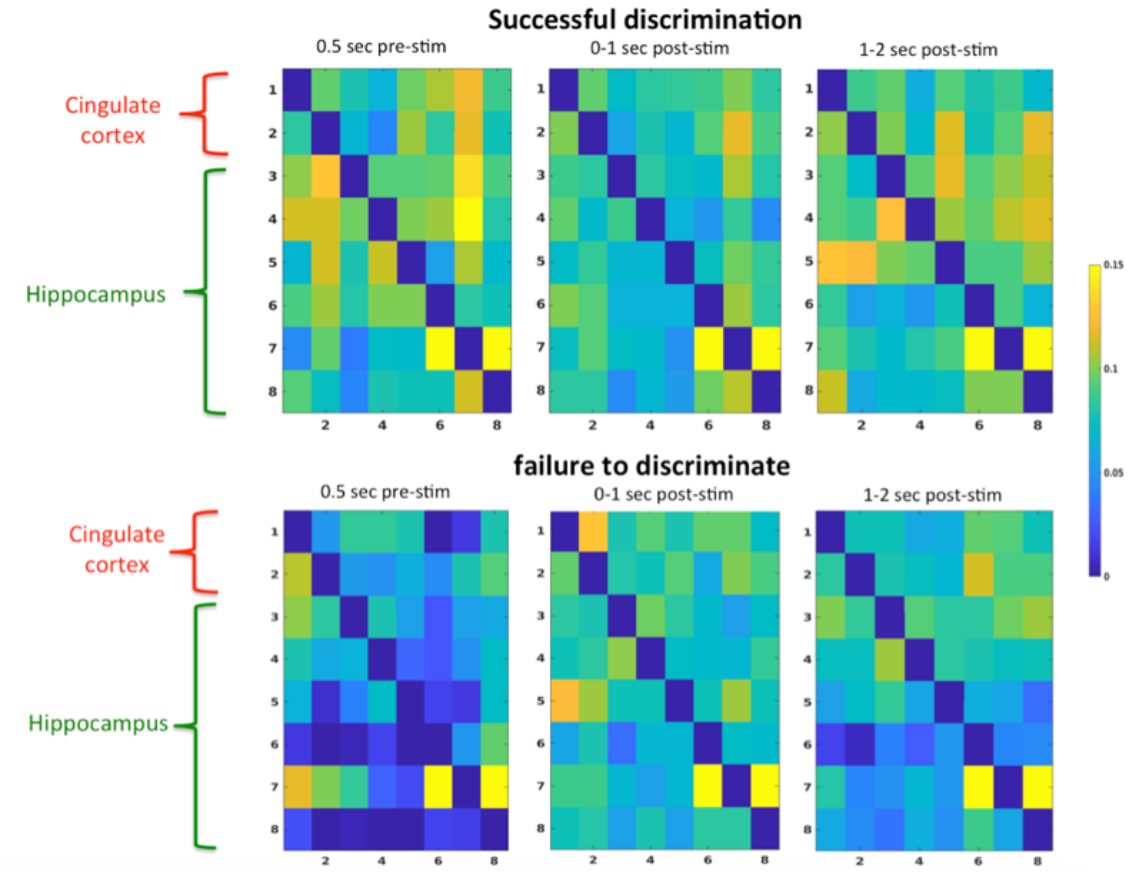


Figure 5a. Neocortical (cingulate cortex)-hippocampal connectivity matrices during a memory task

Mean tv-GCI in the 0.5 seconds preceding stimulus presentation (first column), the first second after stimulus presentation (second column) and the 1-2 seconds after stimulus presentation (third column). This was computed for two behavioral conditions: successful pattern discrimination (first row) and failure to discriminate (second row). Voxel values represent linear dependence (GCI) from row i → column j. Each voxel of the first two rows represents an electrode from cingulate cortex, and the reminder 6 rows represent electrodes from the hippocampus.

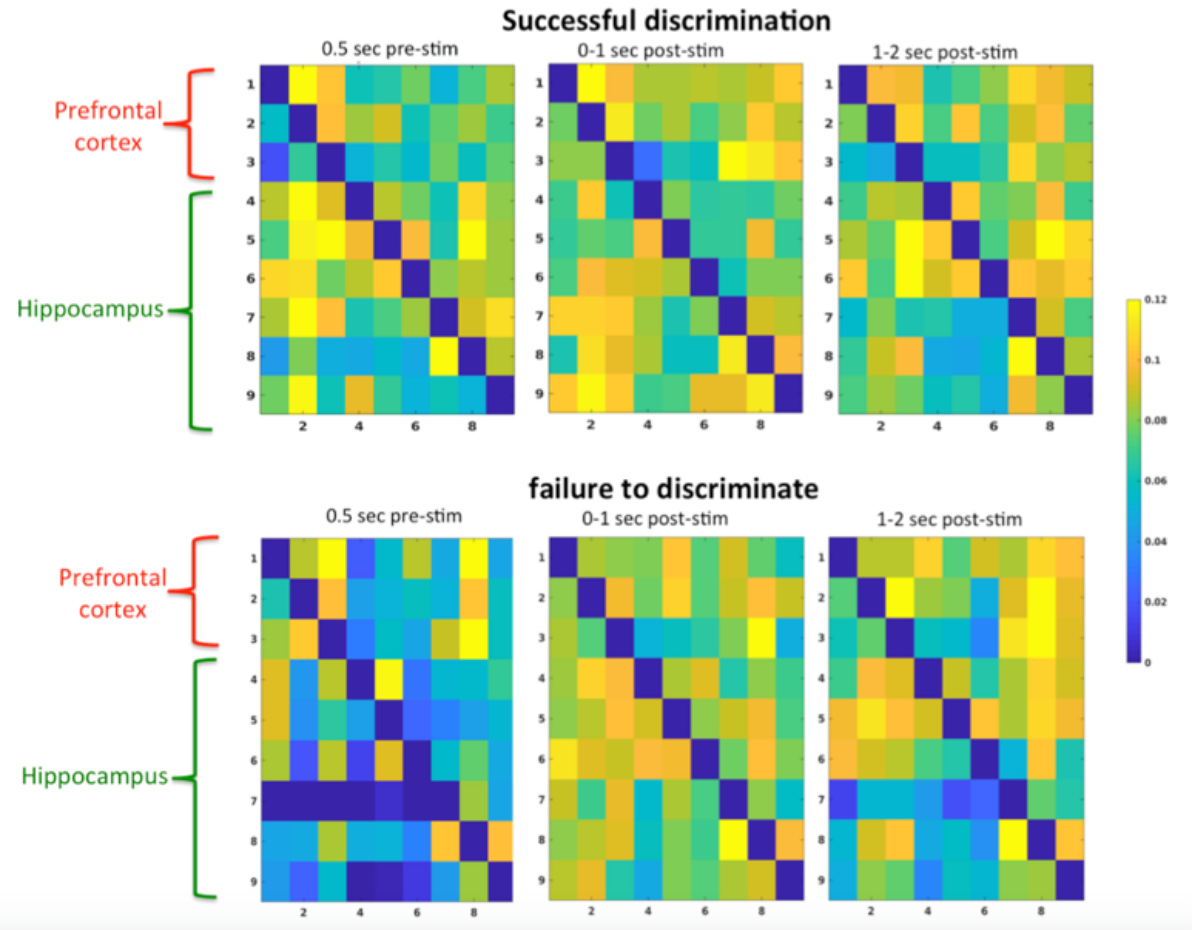


Figure 5b. Neocortical (prefrontal cortex)-hippocampal connectivity matrices during a memory task
Same as in Figure 5a. Here, each voxel of the first three rows represents an electrode from prefrontal cortex, and the remainder 6 rows represent electrodes from the hippocampus.

5.10 An alternative approach, Kalman filter-based granger causality, to estimating directional communication between the hippocampus and the neocortex. Discussion

In this study we utilized the GLKF to fit a TV-MVAR model to intracranial neural data acquired from a human subject during engagement in a memory task. The GLKF is considered to be a useful estimator for this dynamical problem given the desired instantaneous estimates and its improved performance for the multi-trial, multichannel and high SNR case. Moreover, the generation of instantaneous prediction error, which is not the criterion for optimality, makes detection of GCI by means of prediction error ratios feasible.

Using the described methodology, we have shown that the estimated AR model parameters can successfully generate observed data with small RMSE. It is important to note that the number of parameters to data points, in this case, may be too large. Replication of these results with smaller model parameters is crucial to prevent overfitting.

GC analysis revealed hippocampal-neocortical directional influence that supports pattern separation. More specifically, stronger hippocampal influence on neocortical structures is observed only prior to trials that are later successfully discriminated, however, it is not observed prior to failure to discriminate. This could suggest that a minimum baseline communication from the hippocampus to the neocortex is required for successful performance. This baseline communication may involve the initiating of indexing of previously stored memories to extract memory features from neocortex. After this indexing and feature extraction takes place, then successful comparison between previously stored memories and their highly similar currently presented counterparts can take place.

5.11 Kalman filter-based granger causality for estimating directional communication between the hippocampus and the neocortex. References

[1] Arnold M, Miltner WHR, Witte H, Bauer R, Braun C. Adaptive AR modeling of nonstationary time series by means of kaiman filtering. *IEEE Trans Biomed Eng*. 1998;45(5):545-552. doi:10.1109/10.668739.

[2] Ghumare E, Schrooten M, Vandenberghe R, Dupont P. Comparison of different Kalman filter approaches in deriving time svarying connectivity from EEG data. *Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS*. 2015;2015-Novem:2199-2202. doi:10.1109/EMBC.2015.7318827.

[3] Milde T, Leistritz L, Astolfi L, et al. A new Kalman filter approach for the estimation of high-dimensional time-variant multivariate AR models and its application in analysis of laser-evoked brain potentials. *Neuroimage*. 2010;50(3):960-969. doi:10.1016/j.neuroimage.2009.12.110.

[4] Schlögl A, Universit T, Wissenschaften T, Technik B. the Electroencephalogram and the Adaptive Autoregressive Model : Theory and Applications. *Tech Univ Graz*. 2000;Dissertati(April):68. doi:10.1097/MOG.0b013e328333d799.

Chapter 6: System identification of a given brain region reveals comprehensively its basic operational rules (inductive)

Solving the fundamental neural code of the hippocampal CA1 apical dendritic system

6.1 Abstract

Mathematical formulas that fully capture the governing dynamics of synaptic transmission are required for constructing biologically realistic models of brain networks as well as gaining insight into their fundamental operations. We investigated the feasibility of deriving such a formula – an input-output transformation function for the CA3 to CA1 node of the hippocampus. Surprisingly, and in contrast to the field's general belief, *in vitro* slice experiments indicated that these synapses behave in a nonlinear fashion. We used randomly spaced activation of CA3 projections and the Volterra expansion technique for nonlinear system identification of the CA3-CA1 system. The resultant function revealed the key computations of the CA1 (i.e. attenuation, amplification, linear, and nonlinear filtering based on input characteristics). Additionally, the function accurately predicted the continuous waveform of recorded CA1 field excitatory postsynaptic potential (fEPSP) responses with minimal error on data collected from a second set of animals that were not used for function estimation. Thus, the function appears to describe a fundamental and generalizable rule for an essential link in hippocampal circuitry. This work paves the path to uncovering the governing dynamics of other components of the hippocampal system and ultimately, the entire brain.

6.2 Introduction

The hippocampus constitutes a relatively minor (<1%) portion of the human forebrain but plays a pivotal role in the encoding and storage of declarative long-term memories [1]. Theoretical and computational models have long proposed that the dentate gyrus (DG) could

serve to orthogonalize overlapping inputs such that our highly similar day to day memories can be distinctively encoded and stored (i.e., pattern separation) [2-5] Conversely, for memory retrieval, the massive recurrent collateral system of CA3 was hypothesized to act as an auto-associative network, recovering an existing stable pattern when presented with a partial/noisy cue of such memory (i.e., pattern completion) [2-4, 6]. Experimental studies in rodents [7-12] and humans [13-15] have provided convergent evidence that pattern separation and completion are indeed processes instantiated in hippocampal circuitry[16].

These studies collectively use what is, in essence, a 'top-down' deductive strategy in which the goal is to identify local phenomena that can account for macroscopic (behavioral) effects. Conspicuously absent are complementary 'bottom-up' inductive approaches through which identification of fundamental principles can be attained. While the former provides answers to the specific questions tested, the latter reveals physiological rules that link biological phenomena with behavior. Despite significant progress in our understanding of the hippocampus thus far, we are still missing a mapping of its fundamental operations. This has limited our capacity to build a realizable construction of the hippocampal circuit that is complete and accurate, and more generally, to build biological models that are generalizable across problem domains. We therefore employ an inductive approach to comprehensively, agnostically, and rigorously study the hippocampal circuit [17].

System identification, a powerful tool widely used in engineering, aims at fully and accurately characterizing all possible operations performed by an unknown system (i.e., a black box) (**Fig. 1A**). Applying methods from this field to a particular synapse in the hippocampus (an unknown system), we have employed a bottom-up approach to identify nearly all the fundamental principles that govern its dynamics. The Volterra series expansion technique (VSE) for system identification fully captures n^{th} order system dynamics in an agnostic manner (with no prior assumptions about the system, **Fig. 2A right panel**). This agnostic approach prevents the generation of partially complete or misleading models [17]. With sufficiently exciting inputs that

tap into the system's operational range and a clear definition of the hardware of the system being studied (achieved with an *in vitro* setup for establishing causal links between inputs and outputs), a unique solution of a system input-output transformation function – a single formula describing a system that can predict its output to an arbitrary input – is obtained (**Fig. 1A**). We note that the term "transfer function" traditionally refers to linear systems, however, for convenience, our use of this term extends to nonlinear systems (Halás & Kotta, 2007). The power of this single mathematical formulation is not merely the generation of a predictive model, nor is it providing insight into *some* of the tendencies/operations performed by the system. Rather, this function ensures accurate and complete characterization of the entire system under normal conditions which are approximated in our experiments [17]. A consequence of these specifications is a predictive formula that can be applied not just within but *across* animals.

Examination of this transfer function in the time and frequency domains reveals the operations that the system performs. Kernel examination in the time domain reveals the kinds of inputs a given system suppresses or amplifies, how these operations vary as a function of input recency, and the extent of/type of nonlinearities in the system. Examination of kernels in the frequency domain reveals the system's linear filtering properties (1st order kernel) and influence of interactions of multiple frequencies in the input on the frequency content of the output (higher order kernels; i.e. harmonic generation, intermodulation between pairs of input frequencies) [18]. Therefore, the agnostic, complete, and rigorous nature of this approach leads to the discovery of universal mathematical rules guiding the system's synaptic physiology and representing fundamental laws of nature that translate to a major leap forward in our understanding of the brain.

Synaptic transmission relies on currently present input and responses that occurred in the recent past – frequency facilitation and depression are well-studied examples. The factors controlling these interactions are numerous and involve diverse mechanisms operating over different temporal scales (e.g., release dynamics, feedforward IPSCs, summation and shunting,

dendritic channels). Whether the entirety of such immense biological complexity can be accurately and concisely captured in a single mathematical description is an open question. Here, we report the successful derivation of a generalizable transfer function that fully describes up to 2nd order operations performed by CA3:CA1 synapses on arbitrary input patterns in order to generate observed outputs. The function can predict with minimal error the output to arbitrary inputs delivered to the CA3:CA1 node of new animals that were not used for the transfer function estimation. Examination of this transfer function revealed rules about the CA3:CA1 synapses not previously known, to the best of our knowledge, despite this being one of the most heavily studied synapses. Surprisingly, we also find that the dynamics of these synapses violate linearity, which has been a long-standing assumption in the field. These results establish the feasibility of incorporating neurobiologically grounded rules for communication into network models of hippocampus and other structures.

6.3 Result

The CA1 is a nonlinear system

We first tested the prevailing hypothesis that the CA1 is a linear system [12, 16] (**Fig 1B-C**). A system is linear if and only if superposition holds (**Fig 1D**). Superposition principle states that the output to a linear combination of individual inputs would be the same linear combination of the individual outputs. For example, if y_1 is the output to x_1, and y_2 is the output to x_2, then the output to ax_1+bx_2 would be $ay_1+ by_2$ for any weights a, b, and inputs x_1, and x_2. A linear mapping for scalar inputs is a straight diagonal line (**Fig. 1D, bottom plot**) – the long-hypothesized CA1 transfer function. In order to test whether the CA1 is linear, we stimulated the Schaffer collaterals at varying input magnitudes and recorded the corresponding output from the CA1 apical dendrites (**Fig. 1E**). Field excitatory post-synaptic potential (fEPSP) amplitude and slope (**Fig. 1F**) were plotted as scalar values in response to the input (**Fig. 1G**). As shown, these two output measures do not vary linearly with the input, thus, system superposition is negated, and

therefore CA1 apical dendrites fail the linearity test (**Fig. 1F**). To further test whether a linear fit could explain the relationship between the input (stimulation voltage) and output (fEPSP slope), we ran a lack of fit hypothesis test and found that a linear regression line does not fit the data ($p < 0.0001$, $F(79,144) = 74.6$, **Fig. 1G**). This is further evidenced by examining the deviation of the data from the line of best fit; the error is not zero with constant variance. Rather higher error values are present in the earlier and later ranges of x. Noteworthy is that the system behaves in a linear manner over certain ranges of the input (piecewise linear over the approximate x ranges of ~1-4, ~5-9, and ~10-15), when considering slope as a function of the input.

Figure 1

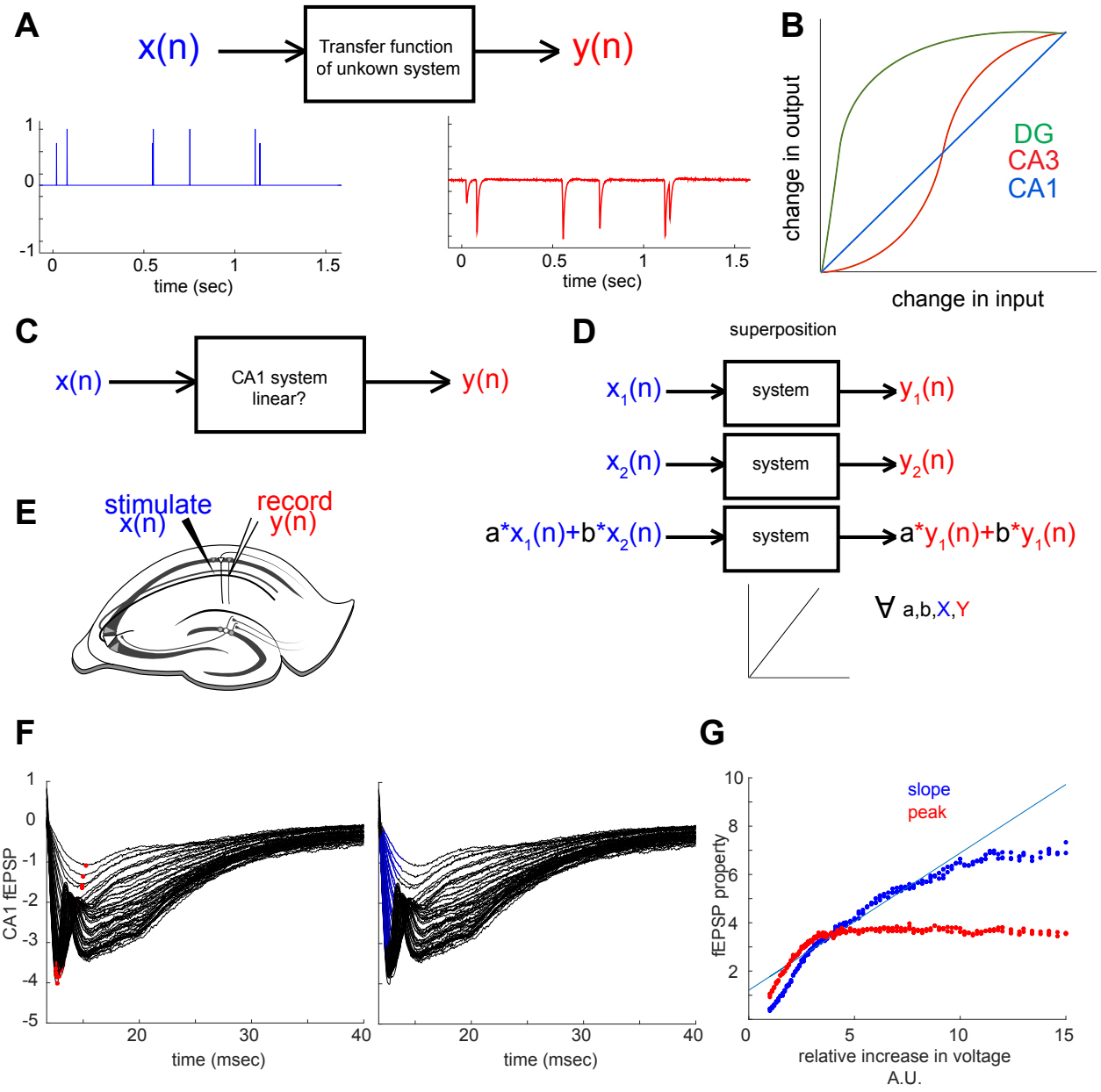


Figure 1. The CA1 is a nonlinear system
A, block diagram. X(n) is the input, y(n) is the output, and the block represents an unknown system. Identifying the function which transforms the input into the output reveals the unknown system. **B,** Hypotheses regarding input-output transfer functions of hippocampal subfields. **C,** CA1 block diagram representation of the popular belief of CA1 linearity. **D,** superposition is a mathematical definition of linearity. Superposition: any linear combinations of any inputs yields the same linear combination of the individual outputs. A linear system must display superposition and will have a line as its input-output curve (bottom). **E,** Schematic of slice electrophysiology recording set up. Inputs X(n) were delivered to the Schaffer collaterals and outputs y(n) were recorded from the CA1 apical dendrites. **F,** Field EPSP peak value (left, red) and slope (right, blue) detection. **G,** CA1 input-output curve. Inputs are varying voltages and

output are fEPSP slope (blue) and peak (red). Line of best fit for slope as a function of input magnitude is overlaid (n=225). Lack of fit test showed that a best fit line does describe the input-output data ($p < 0.0001$, $F(79,144) = 74.6$).

CA1 nonlinear system identification

Characterization of a system's operations (i.e., identifying its transfer function), differs based on whether or not the system is linear. For a linear, time-invariant system, estimating the function *h(i)* would fully characterize the system (**Fig. 2A-B, D**). For a nonlinear system, nonlinearities would not be captured in *h(i)* but in higher order terms (**Fig. 2A-B, E**). Since linearity did not hold in CA1, we identified the CA1 apical dendritic transfer function by estimating both *h(i)* and *h(k,m)* to capture linear and nonlinear components, respectively (**Fig. 2C**), using the VSE technique for system identification [17]. We have truncated this expansion in two ways: 1) system order, p, was defined as 2 whereby only quadratic non-linearities of the system are captured and any higher order nonlinearities that might be present are ignored, and 2) system memory, L, was defined as 60 msec. The term *h(i)*, the first order kernel, is a vector of weights on current and past input values that contribute to the current output (**Fig. 2B, D**). The term *h(k,m)*, the second-order kernel, is represented by a symmetric matrix; its diagonal slices reflect weights for how the multiplicative nature of neighboring input values τ timepoints apart influences the output (**Fig. 2B, E**). Different diagonals reflect weights for different lags (τ's). Within a diagonal (a fixed τ), values further along that diagonal reflect weights for the multiplication of τ neighboring inputs that have already occurred in the more distant past relative to the current output (i.e., the memory of the system, and more simply, the CA1 *record* of its input history, **Fig. 2B**).

Figure 2

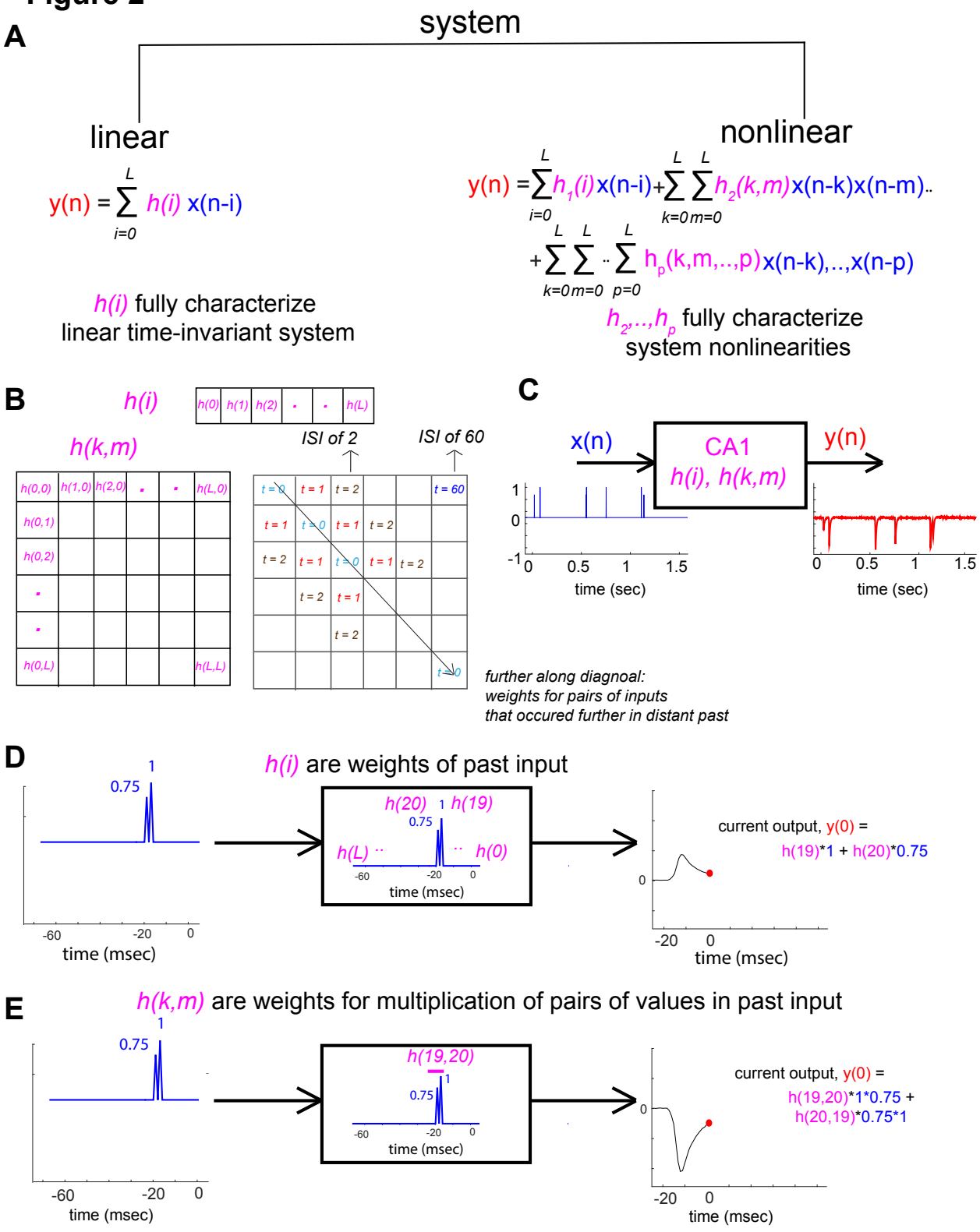


Figure 2. CA1 nonlinear system identification

A, schematic of a division of system types—linear and nonlinear. Corresponding methods to fully characterize each type. **B,** h(i) is a vector of weights. H(k,m) is a 2-d matrix of weights. H(k,m) is a function of two parameters, the inter-spike-interval (ISI, or tau) represented by a fixed diagonal, and how far in the past the ISI occurred (moving along the diagonal) **C,** block diagram for the CA1 non-linear transfer function (h(i), h(k,m)). **D-E,** Schematic representation of how h(i) **(D)** and h(k,m) **(E)** reflect system linear and 2nd order nonlinear operations, respectively. Output traces in D and E are not exact, but approximations to serve as an illustration.

Estimation of first and second order Volterra kernels

To estimate the first and second order kernels, we stimulated the Schaffer collaterals with pulse train signals *x(n)* and recorded the output signals *y(n)* from the apical dendrites of the CA1 (**Fig. 3A**). For each animal (n=11), input signals were delivered in two 15-minute sessions. Input spike time was drawn from a Poisson distribution with a spike rate of 2 spikes/sec and spike amplitude chosen uniformly from two values (x_1 which induces a sub-maximal fEPSP response y_1 of around 3-4 mV, and x_2 as a fraction of x_1 [either ½ or ¾] **Supplementary Figure 2**). These input properties and experiment durations were carefully chosen through simulations to ensure kernel estimation accuracy (**Supplementary Figure 1**).

Similar to other brain areas, the CA1 likely contains subsystems with their own governing dynamics, such as dendritic layer field responses and cell layer spiking (**Fig. 3B**). Linking these subsystems together, a transfer function for input spike patterns *x(n)* to output spike patterns *s(n)* —the fundamental unit of communication in the brain, which is currently uncharacterized—can be obtained. Here, we start with identifying the transfer function (*h(i)* and *h(k,m)*) which transforms spike input into a net apical dendritic field output, $f_{net}(n)$ (**Fig. 3C**). Estimation of these kernels is reformulated as a least-squares regression problem. Group kernel estimates were obtained using the last 12 minutes of the 2nd 15 minute session from each animal and after normalizing the input-output data (**Fig. 3D-E**). The unit impulse response (**Fig. 3E right panel**), which is the sum of the first order kernel (**Fig. 3D**) and second order main diagonal entries (**Fig. 3D-E**), is by definition, a system's response to a delta-function (i.e., a spike). For the dendritic system, as predicted, this is the fEPSP (See **Supplementary Figures**

3-6 for individual animal estimates). The magnitude of the 2nd order kernel diagonal values (1.563) is larger than that of the 1st order kernel (0.6386 **Fig. 3D-E**). Using these two values, the contribution of the linear component (*h(i)*) and nonlinear component (*h(k,m))* to the peak output is quantified as a function of the input (**Fig. 3F, top formula and bottom graph**). The ratio of nonlinear to linear contribution increases with increasing magnitudes of the input (the square of the input is larger, thereby applying *h(k,m)* weights to this larger square contributes more greatly to the output). For the CA1 apical dendritic system, the nonlinear component outweighs the linear one in generating the output at input values larger than ~0.4.

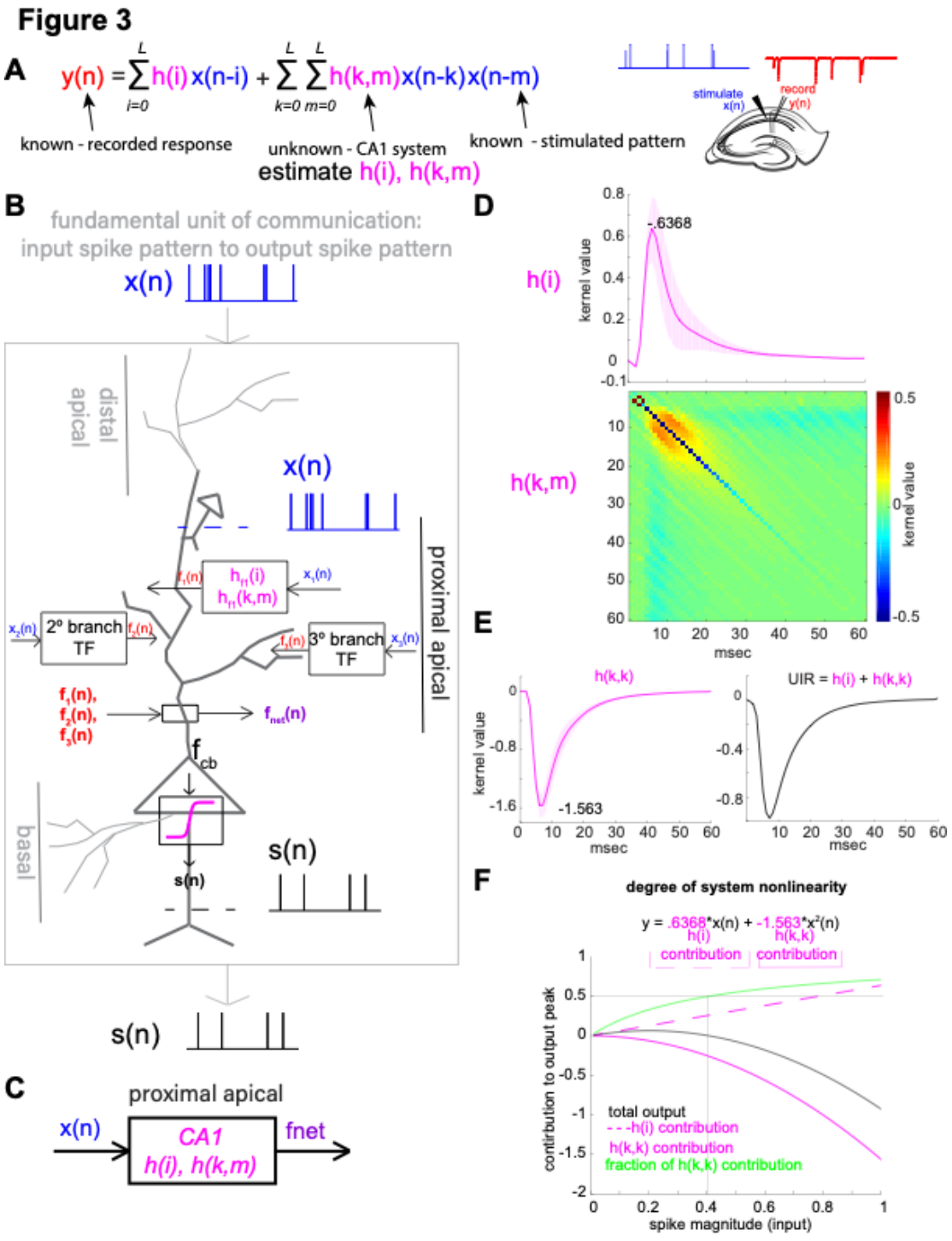


Figure 3. Estimation of first and second order Volterra kernels
A, linking mathematical formulation (left) to experimental design (right). Y(n) and x(n) are the known recorded and stimulated signals, respectively. The CA1 system, h(i) and h(k,m), is unknown. **B,** Large gray box represents the block diagram of input spike patterns x(n) transformed to output spike patterns s(n). Black line focuses on the proximal dendritic domain, and the two other domains are faded (distal apical and basal). Primary branch, secondary and tertiary branches are shown each with their respective inputs, $X_1(n),..,X_3(n)$, field outputs, $f_1(n)$

,..,f_3(n) and nonlinear transfer functions h_{f1}(i), h_{f1}(k,m)… h_{f3}(i), h_{f3}(k,m). f_{net} is the net voltage from dendritic field outputs. This is for a given post-synaptic cell. f_{net} from approximately 15 different cells are measured with the recording set up. f_{cb} reflects the voltage signal at the cell body, which is digitized (sigmoid in magenta) at the axon hillock to 0,1 for spike or no spike, respectively. **C,** Block diagram of the proximal apical dendritic subsystem that transforms Schaffer input x(n) to f_{net} from ~15 cells. **D,** the identified CA1 Apical dendritic system; h(i), h(k,m). **E,** h(k,m) main diagonal slice, h(k,k) (left) and system unit impulse response (UIR) (right). **F,** Quantification of system nonlinearity. Using the peak values of the h(i) and h(k,m) (indicated in E), the predicted output peak (black) is plotted as a function of the input. h(i) and h(k,m) contributions to the output are displayed in magenta dashed and solid lines, respectively. Percent of output generated by the nonlinear system component is shown in green. The intersection of the vertical and horizontal lines indicates that the nonlinear contribution to the output outweighs the linear one for input ranges larger than ~0.4.

Estimated CA1 transfer function reliably predicts CA1 output on new animals

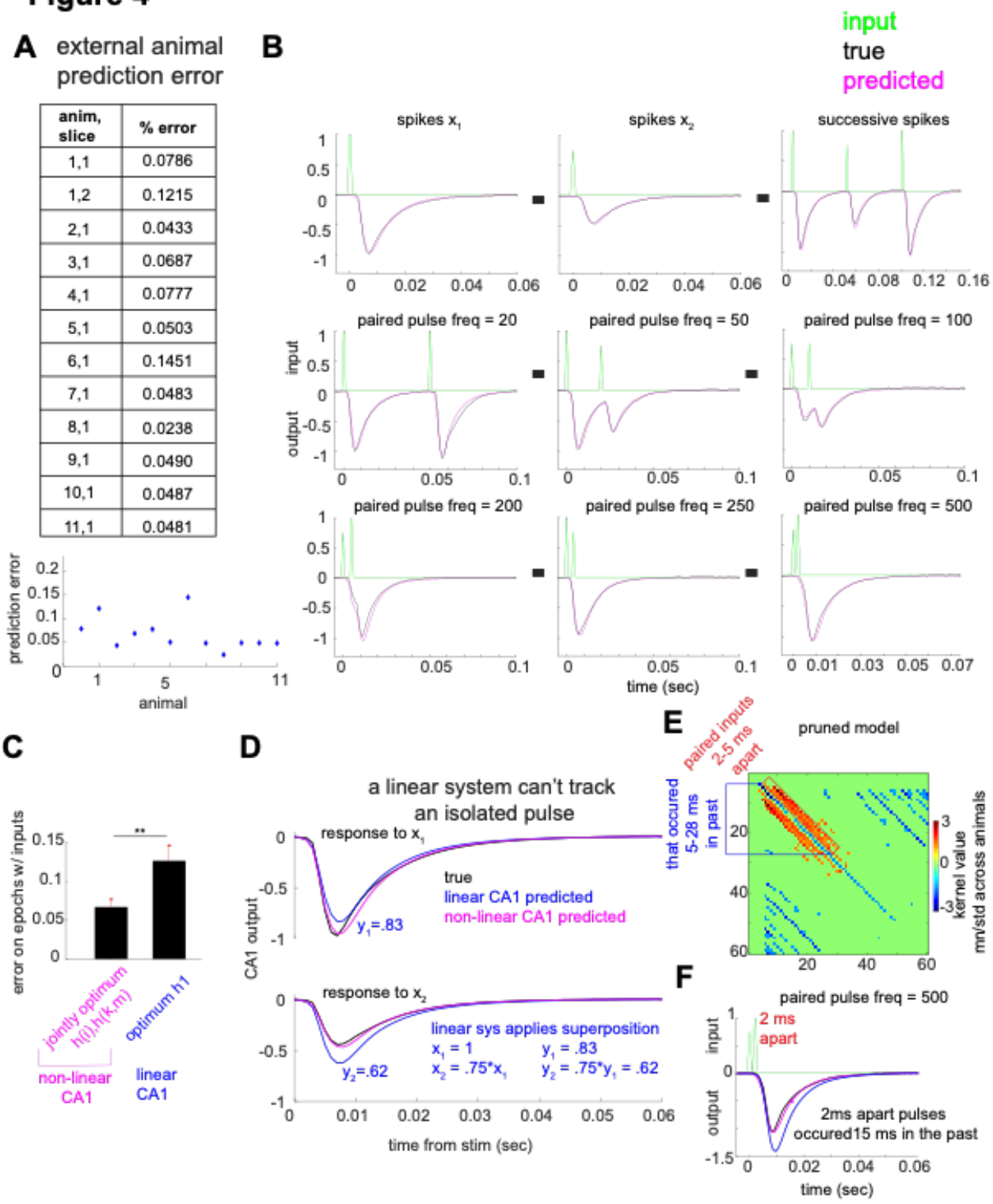
We next asked whether we can use the estimated group transfer function to predict the CA1 response on external data (i.e., in other animals). To do so, 11 different estimates of the CA1 apical dendritic transfer function were made; for each iteration, data from 10 animals were used to estimate the transfer function, which was then used to predict the CA1 output for the 11th animal (**Fig. 4A**). Normalized external prediction error across the 11 animals (12 slices) was 0.0669%. The transfer function was therefore generalizable to new data, and reliably predicted the output, despite the complexity of the input patterns (**Fig. 4B**).

We next quantified the importance of the nonlinear component of the CA1 transfer function to external prediction accuracy, by quantifying the drop in prediction accuracy when using only a first order kernel (linear system) compared to using both the first and second order kernels (**Fig. 4C**). To do so, we separately estimated two transfer functions; one whereby the optimum stand-alone *h(i)* kernel estimates were made, and a second whereby the jointly optimum *h(i)* and *h(k,m)* were co-estimated. We found that prediction error on external data was significantly lower when using the transfer function of a nonlinear CA1 system in contrast to a purely linear system (p = 0.001, paired non-parametric permutation testing).

An estimated linear system failed to predict the CA1 output to both single pulses (**Fig. 4D**) and more complex temporal pulse patterns (**Fig. 4E**). Using single pulse inputs, a linear

system failed to predict the change in output fEPSP amplitude with a change in input stimulation – since it assumes superposition (i.e., increasing the input by 75%would proportionally increase the fEPSP negative peak by 75%, **Fig. 4D**). This superposition is also the only mechanism employed by a linear system to predict the output on successive pulses at random times and amplitudes, leading to error accumulation (**Fig. 4F**). Conversely, the estimated nonlinear transfer function tracks isolated inputs of variable magnitudes (**Fig. 4D**) as well as arbitrary pulse patterns (**Fig. 4B**) by accounting for their time of occurrence and the lag between all possible pairs of spikes (**Fig. 4F**). Examination of the second order kernel weights reveals that the CA1 amplifies consecutive pulses that are 2 to 5 msec apart (leading to a less negative fEPSP), having occurred anywhere from 5 to 30 msec in the past relative to the current time point being predicted (**Fig. 4E inset**). These ranges were extracted from a pruned second order kernel. Included in this pruned matrix are the entries of the second order kernel whose values were either below or above the 5^{th} and 95^{th} percentiles, respectively, of the mean divided by the standard deviation of entries across all animals (consistently large negative and positive weights, respectively, across animals). Capturing these CA1 dynamics in the second order kernel enables the nonlinear transfer function to reliably predict the output to a random input.

Figure 4



anim, slice	% error
1,1	0.0786
1,2	0.1215
2,1	0.0433
3,1	0.0687
4,1	0.0777
5,1	0.0503
6,1	0.1451
7,1	0.0483
8,1	0.0238
9,1	0.0490
10,1	0.0487
11,1	0.0481

Figure 4. Estimated CA1 transfer function reliably predicts CA1 output on new animals

A, Table (top) and (bottom) plot displaying external animal prediction error (model generated using n-1 animals, with the test animal excluded). **B,** Example of predictions from animal 8 data for isolated spikes (top row, first two panels), successive spikes (top row, third panel), and paired pulses of variable inter-spike-intervals ISI (remaining). All examples were extracted from a continuous 15-min recording. **C,** Differences in external animal prediction error using jointly optimum h(i), h(k,m) and the optimum h(i), reflecting nonlinear and linear systems, respectively (p = 0.001, paired non-parametric permutation testing). **D,** Black traces reflect mean true response to x_1 (top) and x2 (bottom) to isolated spike inputs (not preceded nor followed by spikes in the prior 1-sec and 0.25 sec, respectively). Predicted waveforms by linear and nonlinear CA1 are shown in blue and magenta, respectively. Linear CA1 applies superposition (y2 is the same linear combination of y1 as x2 is of x1.). **E,** Pruned h(k,m) displays a thresholded matrix where included voxels are those in the top and bottom 5% of the standardized mean (mean/standard deviation) kernel value distribution (top). Positive weights (red rectangle) are applied for input pairs with an ISI of 2-5 ms. Weights attenuate as a function of the ISI past (blue bracket). Negative weights are present for ISI's of 40-60 ms (blue voxels). **F,** Illustration of utility of pruned model for output generation. An input of two spikes that are 2 ms apart is shown in green. Part of the predicted output value at 15ms, generated by the nonlinear term, is the input weighted by the 15^{th} element of the matrix slice that is 1-off the diagonal.

6.4 Discussion

The work presented here is a formal identification of the computational operations of the hippocampal CA1 apical dendritic system. We first experimentally tested the general belief that the transfer function for the Schaffer collateral input to CA1 output is a linear function [12, 16, 19-21] and found clear evidence for non-linearity (**Fig. 1D**). In further support of this finding, we demonstrate that if the CA1 transfer function is estimated based on assumed linear dynamics, such transfer function fails to predict not only the output to complicated input patterns, but also responses elicited by single pulses of variable amplitude (**Fig. 4D**). To identify the nonlinear transfer function of the CA1 subsystem, we used the VSE – an agnostic technique that completely characterizes an r^{th} order system with r^{th} order kernels, capturing the entirety of a system's dynamics, and enabling the prediction of output in response to arbitrary input [17]. We reliably identified the first and second order kernels in individual animals (**Supplementary Figure 3-4**) and, using this identified CA1 subsystem transfer function, successfully predicted with high accuracy the dendritic output to complex input patterns delivered in a separate group of animals (**Fig. 4**).

Examining this transfer function revealed new insight about system dynamics. First, the summation of the first order kernel and the diagonal entries of the second order kernel, by definition, are the unit impulse response (UIR) of the system (response of a system to activation of its afferents with a spike input) (**Fig. 3D-E**). Indeed, the sum of these two estimates yields the fEPSP (**Fig. 3E right panel**).

Second, we found that deviations of the system dynamics from linearity can be quantified in a generalizable manner and are dictated by the ratio of *h(k,m)* to *h(i)* magnitude. The larger the ratio, the earlier the deviation occurs (at smaller values of *x*). For the CA1 apical dendritic system, nonlinearities make up more than 50% of output magnitude at normalized input values larger that ~0.4; at these values, the 2nd order kernel contributes more to the fEPSP waveform. More generally, this approach of quantifying system nonlinearity provides a means for the canonical comparisons of nonlinearities across the hippocampal subfields (DG, CA3, and CA1 **Fig. 1B**), which is a key rationale for subfield engagement in the core memory processes of pattern separation and completion.

Third, use of the 2nd order kernel was critical to capturing system 2nd order nonlinearities. More specifically, in the time domain, *h(k,m)* contains a weight for the multiplicative interaction of all possible pairs of input values and their distance in time from the present output. This allows for temporal summation to occur, whereby the influence of the multiplicative interaction of neighboring values of the input and CA1's memory for such neighboring values, dictate the output. Examination of the CA1 apical dendritic system 2nd order kernel reveals that the positive weights (less negative fEPSP) are applied to neighboring values of the input that are ~2-5 msec apart (paired pulses in the ~150-500 Hz range), keeping them in memory for ~5-28 msec. In contrast, the system applies negative weights (more negative fEPSPs) to neighboring values of the input that are ~40-60 msec apart (paired pulses in the ~16-25 Hz range) (**Fig. 4E-F**). The comprehensive approach used in the present work has thus revealed previously undetected physiological operations in one of the most intensively studied synapses. Notably, unexpected

rules afford an opportunity for strong synergies between the current inductive strategy and the deductive investigations conventionally used in neuroscience. This would involve the development of *a priori* hypotheses about how physiological processes with different time constants (e.g., transmitter mobilization, feedforward inhibition) interact to produce generalizable rules.

While kernel estimates were relatively consistent from animal to animal, the first order kernels did show some variability (**Supplementary Figure 3**). This is due to the presence of two types of first order kernels extracted across animals; one which resembled an upward sloping fEPSP (type #1; reflected in the grand average, **Fig. 3D**) and a downward sloping one (type #2). This difference in 1st order kernel estimate is consequential and informs as to the nature of system output as a function of input. 1st order kernel type #2 will lead to a *more negative* fEPSP contribution in addition to the negativity generated by the 2nd order kernel. This greater negativity effect will be even larger for smaller inputs where *h(i)* predominates the response compared to *h(k,m)* (compare **Supplementary Figure 2-3**).

The low amplitude fEPSPs used in the present study are generated by a small fraction of the synapses (~15) at the recording site in the proximal apical dendrites of CA1. Termination patterns, and thus spatiotemporal summation of EPSPs, will likely differ significantly across stimulation and recording electrode sites (**Fig. 3B**). Similarly, the number of contacts with feedforward interneurons that shunt excitatory currents in pyramidal cells will differ between any two small collections of stimulated axons. Together these factors provide a reasonable explanation for the observed variation in *h(i)* kernels.

The synapse represents a fundamental unit of communication in the brain – where input spike patterns are transformed into output spike patterns (**Fig 3B**). Knowing the transfer function for this process is essential to generating an adequate theory for the brain. This transformative process involves sequential intermediary transfer functions whereby dendritic field response $f_{net}(n)$ to a sequence of afferent action potentials $x(n)$ is first generated and followed by spiking

s(n) by the target cells (**Fig 3B**). Results reported here describe a transfer function for the dendritic component of the system while earlier system identification work provided insights regarding voltage fluctuations at the cell body (f_{cb} **Fig, 3B**, Cook et al., 2007) and translation into spike output [22-27].

Included in the latter are quantitative analyses of voltage transfer along the main dendritic shaft to the f_{cb} [28], information that is crucial to modeling EPSP-initiated spiking patterns. Other studies have used system identification to extract relationships between random interval afferent stimulation and the production of synchronous spike discharges by multiple neurons (population spikes) in the three nodes (dentate gyrus, CA3 and CA1) of the trisynaptic hippocampal circuit [22-24]. While transfer functions for single event (pop-spike) results reveal some of the system's properties, they do not bear directly on *patterns* of output spikes, which is required for full characterization of the fundamental unit of communication in the brain.

Systematic comparisons of input-output spike patterns have been conducted using *in vivo* recordings from CA3 (input stage) and CA1 (output stage) during well-defined behaviors [25-27]. By design, such studies utilize a restricted set of input patterns (those associated with a particular context) and thus cannot produce the generalized kernels obtained with sufficiently exciting inputs in an *in vitro* model. Moreover, correlations between the two subfields *in vivo* could in part reflect near simultaneous input to both regions from a third region such as entorhinal cortex or septum, areas known to synchronize activity within the hippocampus. Nonetheless, the results provide special cases that must be satisfied by universal models. Constraints are also provided by prior work reducing analog outputs to univariate measures (fEPSP slope and pop-spike magnitude) of CA1 elicited by specialized inputs (fixed frequency stimulation) recruiting different numbers of Schaffer collateral axons [29]. Thus, while prior work has generated important insights about input/output relationships throughout the trisynaptic circuit, the current work extends these insights to produce generalizable and fundamental rules guiding the operations of a well-studied system, the Schaffer collateral to CA1 synapse. While

this work identifies the transfer function of input spike patterns to the CA1 apical dendritic field (**Fig. 3C**), further work is needed to link concise mathematical descriptions of the CA1apical dendritic field to output spiking patterns by pyramidal neurons.

There are a few limitations to the current work. First, two assumptions were made in implementation of system identification in the present study. The CA apical dendritic system was defined to be 2nd order with a 60 msec memory – these definitions captured the fEPSP transformations with small error (93% external animal prediction accuracy). Through a qualitative examination, we found that the 60 msec assumption is a key contributor to prediction error. Consistent with prior work, it is likely that the CA1 keeps a longer record of the input [29]. There is also a potential for higher than 2nd order dynamics [24], although this was not apparent in our study through a qualitative examination of model predictions to isolated paired pulses (a system is second order if a second order model can predict accurately the responses to isolated paired pulses). Future work is needed to formally test system memory length and order. Second, VSE is a nonlinear system identification technique for time invariant systems (system dynamics are stationary as a function of time). We indexed and estimated kernels using the most stationary segment in the data (12-min, see methods). However, our results do not address instances in which system dynamics change quickly and dramatically as in the case of long-term potentiation or depression. Whether system dynamics as captured in Volterra kernels are altered under these circumstances is an intriguing and fundamentally important future question to address.

Other connections in the hippocampal network differ in important ways from the CA3:CA1 link studied here and have diverse types of specializations. They are likely to have transfer functions that are quite different from that described above. It will be of considerable interest to determine if these bear any logical relationship to inferred functions such as pattern separation and completion. Work in this area will also serve to evaluate the feasibility of integrating the transfer functions for multiple sites into a large-scale model for the entire

hippocampal network. Such an enterprise would first require tests of whether the strategy used in the present work to define dendritic responses to diverse patterns of input can be applied to the spiking patterns triggered by the responses (**Fig. 3B**). If so, then it should be possible to create a robust and realistic hippocampal simulation and ask if operations that emerge naturally from it can be directly linked to observed behavioral phenomena. Connecting this to extant cortical models, a step that will be facilitated by the economical nature of transfer functions, will be a quantum leap towards brain-based artificial intelligence.

6.5 Methods

Animals

Experiments were conducted using 11 male mice (10 C57BL/6N and 1 sighted FVB129) 2-3 months of age (**Supplementary Table 1**). Animals were group-housed (4-5 per cage), on a 12-hr light/dark cycle, and with food and water *ad libitum*. All procedures were approved by the UC Irvine Institutional Animal Care and Use Committee and in line with the NIH guidelines for the Care and Use of Laboratory Animals.

Hippocampal slice electrophysiology

Animals were sacrificed between 10:00am-11:00am. Acute hippocampal slices were prepared as described previously [30]. Briefly, transverse slices of the middle third of the left hippocampus (360 um thickness) were cut using a McIlwain chopper and collected in chilled artificial CSF (ACSF) which contained in mM (124 NaCl, 3 KCl, 1.25 KH_2PO_4, 1.5 $MgSO_4$, 26 $NaHCO_3$, 2.5 $CaCl_2$, and 10 dextrose). Slices were then transferred to the interface chamber kept at 31±1°C and had 60-70 mL/hr infusion of oxygenated ACSF. The interface chamber was prepared an hour prior to slice preparation. Slices were left for 1.5-2 hours to equilibrate prior to recording. A stimulating electrode (twisted nichrome) was placed in the CA3-CA1 Schaffer collaterals and a glass recording electrode (2 M NaCl filled, 2-3 MΩ) was used to record

extracellular field potentials from the CA1 apical dendrites. The stimulating and recording electrodes were placed in the CA1b stratum radiatum with the stimulating electrode slightly offset towards the stratum pyramidale, and the recording electrode in the middle of the stratum radiatum [30]. In most animals, stimulation intensity (x_1) was set to induce a ~3.5-4 mV population-spike free fEPSP (**Supplementary Figure 2**).

Recording protocol

For each animal, a 10-minute stable baseline session was recorded before experiment initiation. Baseline stimulation consisted of a single pulse every 20 seconds. After ensuring stable baseline responses, two 15-min sessions were recorded per animal. Slices were left to stabilize for 5 minutes without stimulation between the two 15-min sessions. Stimulation patterns were random with spike times drawn from a Poisson distribution (lambda of 2 spikes/second) and amplitudes uniformly drawn from two values (x_1 and x_2: x_1 induced a 3-4 mV fEPSP and x_2 was a fraction of x_1**, Supplementary Table 1**). Recording duration and patterns were explored and determined through simulations such that they yielded unbiased parameter estimates (**Supplementary Figure 1**).

Equipment

The stimulation signal *x(n)* was generated digitally in MATLAB 2018a (The MathWorks, Natick, MA), converted to an analog signal using a 16-channel data acquisition hardware (DA-16, A-M Systems), which was then connected to a stimulus isolation unit (2200, A-M Systems) enabling the delivery of current to the stimulating electrode. DA-16 was also used to digitally record the delivered analog input *x(t)* and CA1 analog output *y(t)*, sampled at 10,000 Hz and saved in MATLAB. Both stimulus delivery and recording implemented through the DA-16 were interfaced using MATLAB and its Data Acquisition Toolbox for Measurement Computing.

Analysis: Simulations

Prior to initiation of data collection, simulations were conducted to aid in the appropriate experimental design for accurate and unbiased Volterra kernel estimates. Ground truth first and second order Volterra kernels were generated, using knowledge of the structure of these kernels (symmetric second order kernel) and candidate CA1 dynamics (**Supplementary Figure 1A**). The ground truth first order kernel was defined as an fEPSP. This is because h(i), for a linear system, is the response to a delta or spike input and the CA1 apical dendritic field generates an fEPSP to a spike input. Using these ground truth kernels, we tested the nature of the input and amount of data needed for accurate kernel estimation.

First, we found that increased randomness in the input increases the likelihood of tapping into the second order structure – with such input kernel weights for all possible tau's are applied and reflected in the output, such that they can be recovered with the input/output data. We therefore decided to deliver inputs at random times pulled from a Poisson distribution, which also mimicked the underlying distribution of neural spike timing (**Supplementary Figure 1B**). The Poisson parameter lambda was set to 2 spikes/second. A slow spike rate was chosen to prevent overstimulating the slice thereby affecting its dynamics during the recording, and this spike rate was used in prior work [31-33].

Second, we found that a binary input leads to estimation bias due to multicollinearity – where the first order kernel and the diagonal elements of the second order kernel become coupled. This is because the first order kernel provides weights for individual input values and the diagonal of the second order kernel provides weights for the square of such values (**Fig. 2B, D-E**). With a binary input (0 and a non-zero value), the square of the input vector becomes a linear combination of it leading to multicollinearity and estimation bias. We therefore randomly drew the spike amplitude uniformly from two values, x_1 and x_2 (x_1 induced a ~3-4 mV, and x_2 was a fraction of x_1 either ½ or ¾, see **Supplementary Table 1** and **Supplementary Figure 2**).

Finally, for the desired model order – a 60 msec kernel length and up to a second order kernel, we quantified normalized estimation error as a function of data duration (**Supplementary Figure 1 C-D**). The error reached a plateau of 0 with 13-15 minutes for a single session.

Analysis: Data curation and preprocessing

Data curation and preprocessing were done offline in MATLAB. The recorded input signal was used to identify the input timestamps. The detected timestamps were then used to remove stimulation artifacts from the CA1 record. Local points (17 data points ~1.7 ms) around the artifact peak were replaced by propagating the CA1 signal immediately before the artifact using the slope of the signal at the boundaries of the artifact (before and after) plus noise pulled from a normal distribution with 0 mean and the standard deviation of the CA1 artifact free record weighted by 1/16. The data was then downsampled to 1000 Hz. All individual animal kernel estimates took place subsequently using this data (non-normalized).

For group analysis, additional preprocessing was implemented and consisted of within animal 1) pre-stimulation normalization and 2) normalization of fEPSP amplitude relative to the mean fEPSP induced by x_1. Pre-stimulation normalization accounts for slow drifts in the data, which offset the starting value of the induced fEPSP. Each value following a stimulus was z-score normalized relative to the mean and standard deviation of a 100 msec (1000 data points, before downsampling) period prior to stimulation that does not overlap with a CA1 response (up to 80 msec) from a prior stimulus. For successive stimulations, however, this baseline period did not exist. Therefore, successive stimuli occurring proximally in time shared the same pre-stimulus baseline period which occurred before the train.

After pre-stimulus baseline normalization, all values were normalized to the negative peak of the mean fEPSP trace induced by x_1 (~3-4 mV). This enabled all $y(n)$ output values to be bounded between 0 and ~-1 for all animals. Additionally, this enabled normalization of the

input signal across animals; whereby the x_1 inducing ~ -1 normalized response was arbitrarily mapped to 1 for all animals. Then, x_2 became simply equal to the original fraction used during the experiment – either ½ or ¾.

Analysis: Kernel estimation

Out of two 15-min sessions recorded per animal, the last 12-min data (most stationary subset) of the 2nd 15-min session was utilized to obtain individual animal estimates for subsequent group analysis. This choice accounted for the tradeoff between stationarity and sufficient data for accurate estimation (**Supplementary Figures 1, 3-4**).

The curated CA1 output signal was used as the *y(n)* vector, and an equal length spike vector, *x(n)*, was re-arranged into a matrix, X_{design} such that the kernel estimation is reformulated into a least squares regression problem $y(n) = X_{design} * h_{all}(n)$ and solved using the pinv function in MATLAB. $h_{all}(n)$ contains all kernels (both first and second order) in vector format. It is important to note that in generating the design matrix, we took advantage of the symmetry of the second order kernel and only estimated the diagonal and upper triangular elements of the kernel to reduce the number of unknown parameters.

Analysis: normalized prediction error

Normalized prediction error was quantified as the sum of the squared error (predicted - true) normalized by the sum of the squared true output [22]. This measure was calculated for stimulation epochs (each stimulation time point and the following 60msec data). Such epochs are ones where non-zero predictions were made, otherwise the system is assumed to be at rest.

$$normalized\ prediction\ error = \sum_{\substack{stim\ epoch \\ timepoints}} \frac{(y_{predict} - y_{true})^2}{{y_{true}}^2}$$

Analysis: Statistical analyses

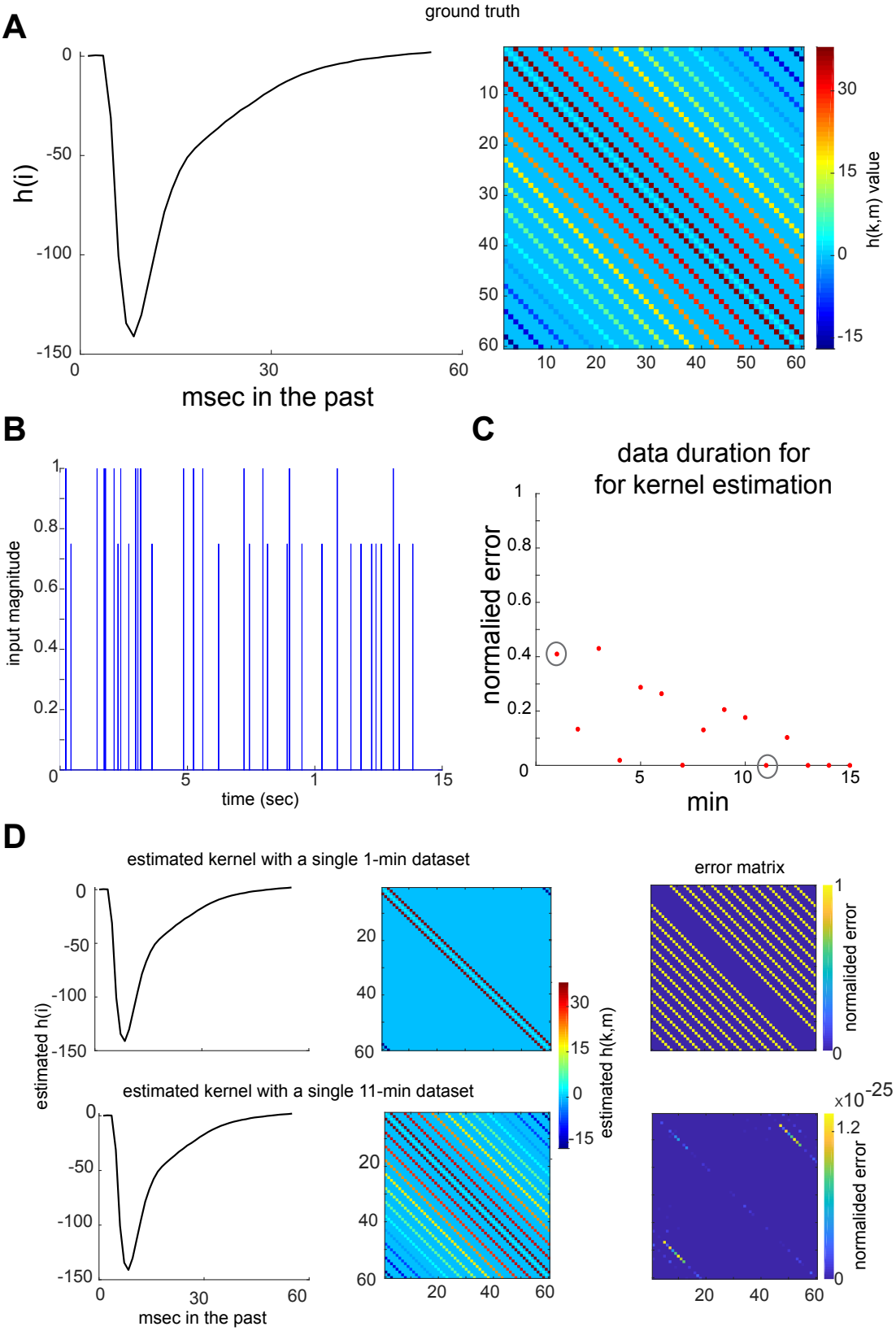
Significance testing on prediction error was done using non-parametric paired permutation testing with 100 permutations.

6.6 Supplementary material

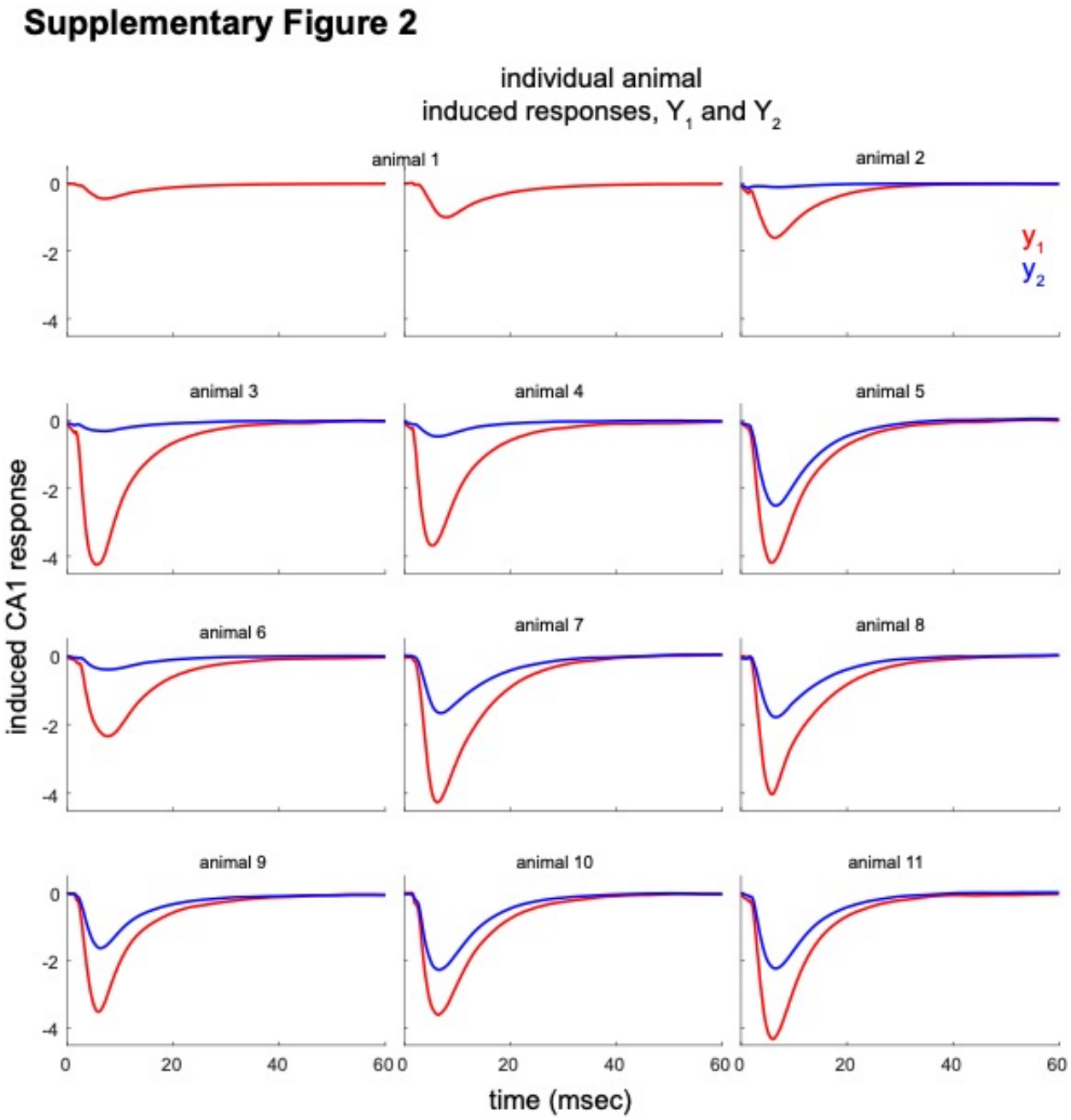
animal #	# of slices	# of trials	Trial duration	Stim protocol
1	2	3,3	20	X_1
2	1	2	15	X_1, ½*X_1
3	1	1	15	X_1, ½*X_1
4	1	2	15	X_1, ½*X_1
5	1	2	15	X_1, ¾*X_1
6	1	2	15	X_1, ¾*X_1
7	1	2	15	X_1, ¾*X_1
8	1	2	15	X_1, ¾*X_1
9	1	2	15	X_1, ¾*X_1
10	1	3	11,15,15	X_1, ¾*X_1
11	1	2	15	X_1, ¾*X_1

Table 1.
Table displaying recording information regarding each animal.

Supplementary Figure 1

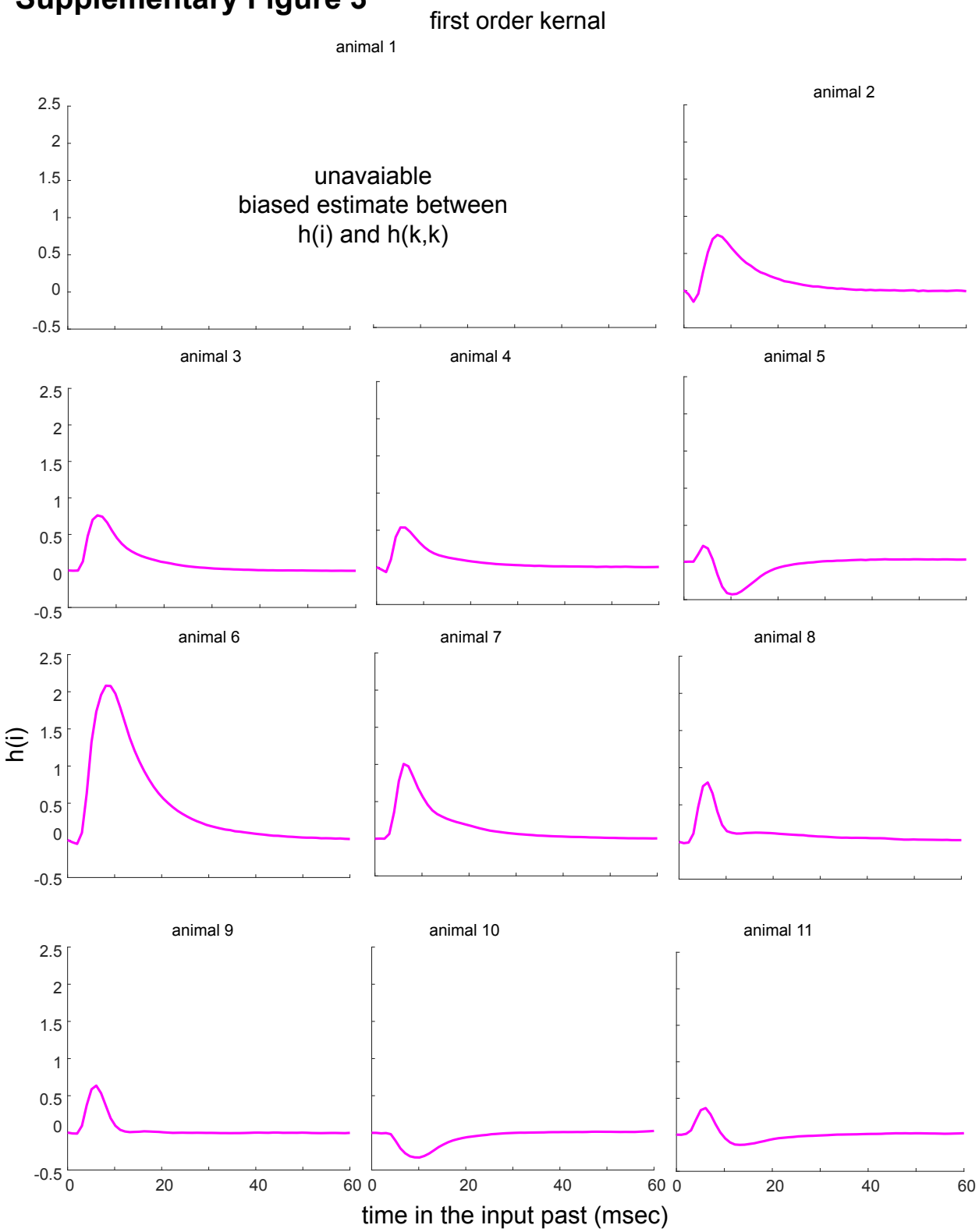


Supplementary Figure 1.
Simulation to identify recording duration and input type needed for accurate and unbiased kernel estimation. **A,** Artificially generated ground truth first (left) and second (right) order kernels. **B,** example of nonbinary (0, X_1 =1 or X_2 = 0.75* X_1) input with ISI from a Poisson distribution and spike amplitude from a uniform distribution. **C,** Normalized estimation error as a function of data duration. **D,** Example kernel estimates (left) and h(k,m) error matrix (right) using 1 minute (top) and 11 min (bottom) data, whose total errors are circled in **C.**



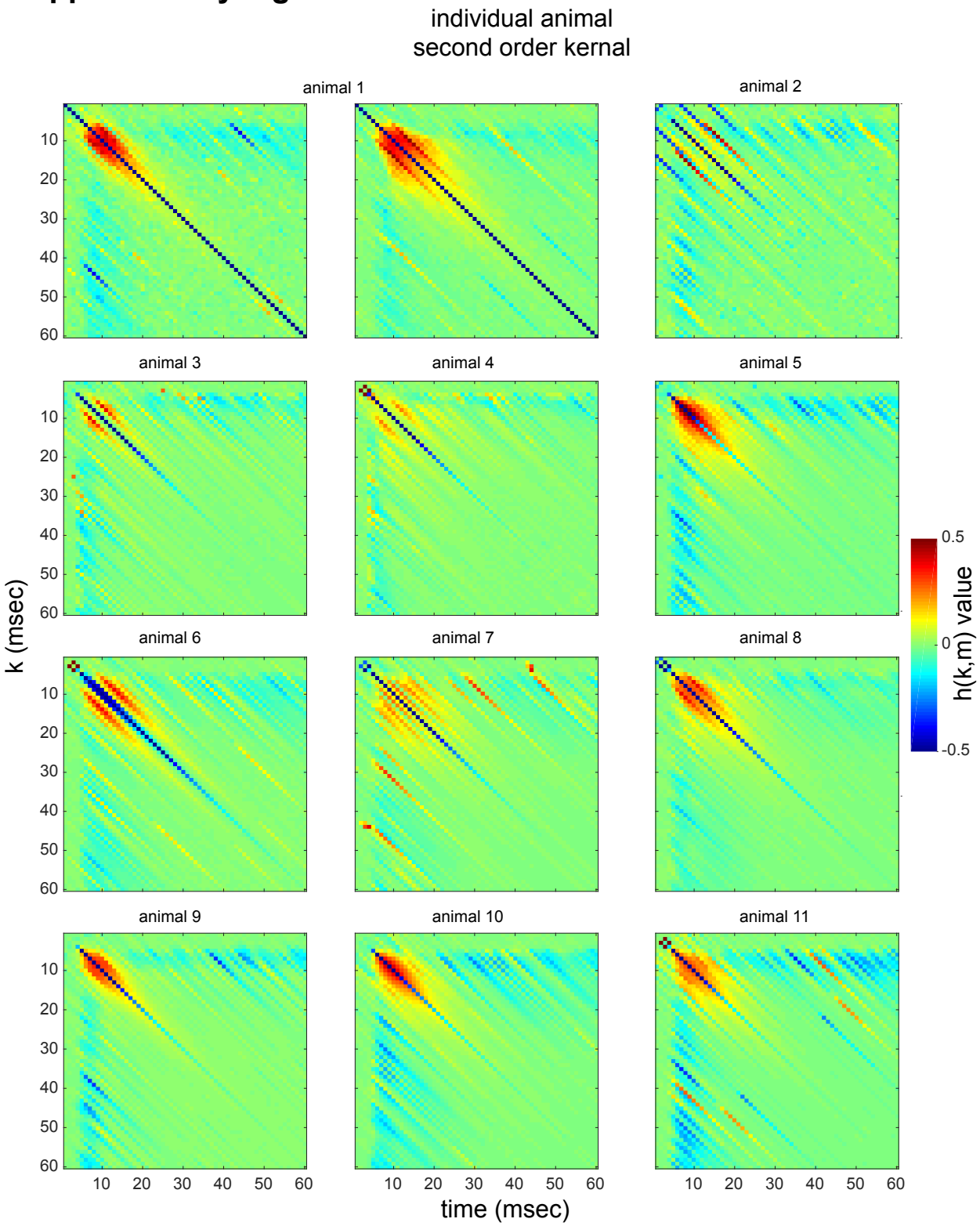
Supplementary Figure 2.
Individual animal induced mean responses Y_1 and Y_2.

Supplementary Figure 3



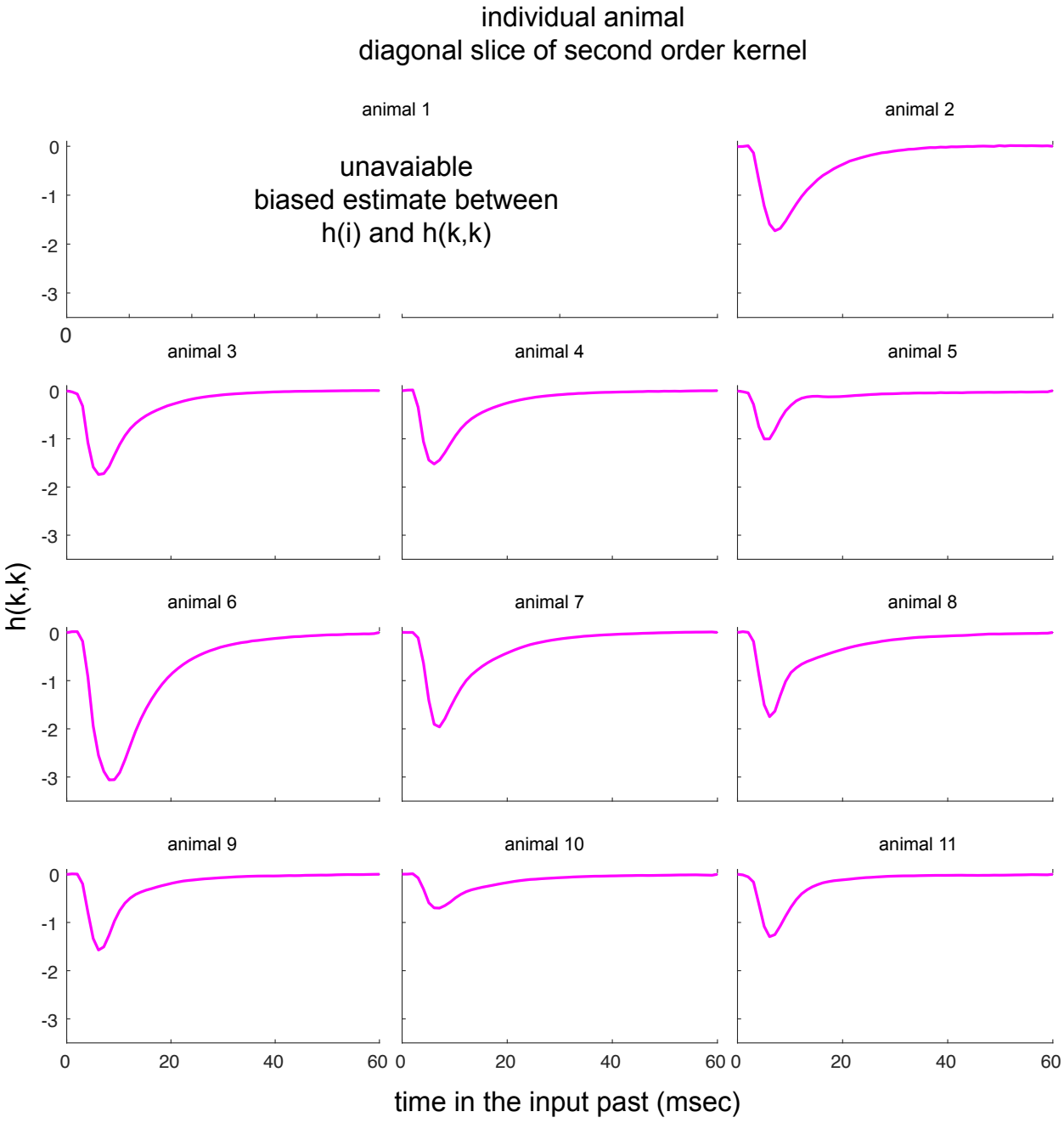
Supplementary Figure 3.
Individual animal h(i) estimates using the last recorded 12-minutes for each animal.

Supplementary Figure 4



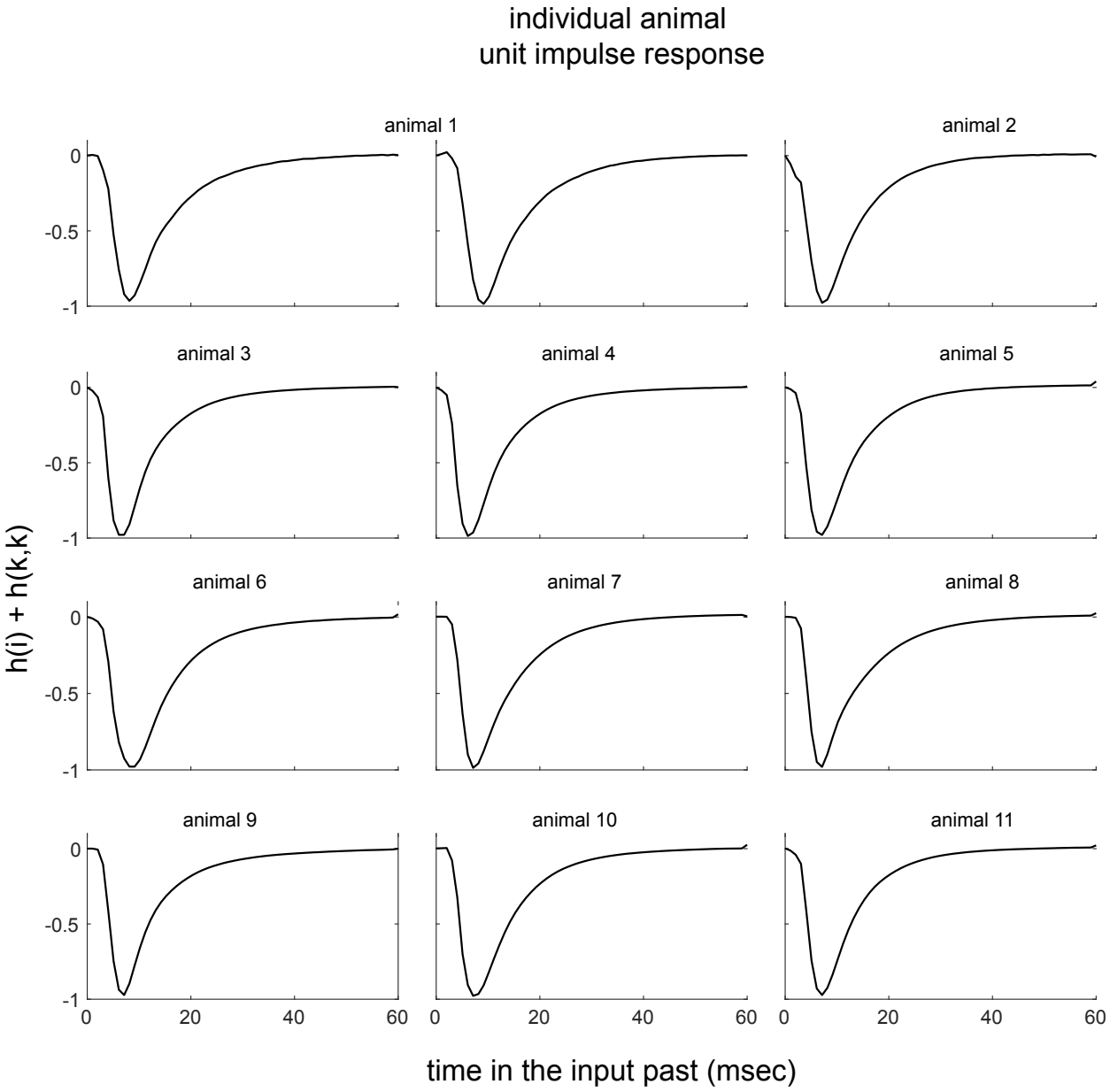
Supplementary Figure 4.
Individual animal h(k,m) estimates using the last recorded 12-minutes for each animal.

Supplementary Figure 5



Supplementary Figure 5.
Individual animal diagonal slice h(k,k) of the second order matrix h(k,m).

Supplementary Figure 6



Supplementary Figure 6.
Individual animal unit impulse response (UIR).

6.7 References

1. Squire, L.R., C.E. Stark, and R.E. Clark, *The medial temporal lobe.* Annu Rev Neurosci, 2004. **27**: p. 279-306.
2. O'Reilly, R.C. and J.L. McClelland, *Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off.* Hippocampus, 1994. **4**(6): p. 661-82.
3. Rolls, E.T., *A computational theory of episodic memory formation in the hippocampus.* Behav Brain Res, 2010. **215**(2): p. 180-96.

4. McClelland, J.L., B.L. McNaughton, and R.C. O'Reilly, *Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory.* Psychol Rev, 1995. **102**(3): p. 419-457.
5. Marr, D., *Simple memory: a theory for archicortex.* Philos Trans R Soc Lond B Biol Sci, 1971. **262**(841): p. 23-81.
6. Rolls, E.T., *Pattern separation, completion, and categorisation in the hippocampus and neocortex.* Neurobiol Learn Mem, 2016. **129**: p. 4-28.
7. Leutgeb, S., et al., *Distinct ensemble codes in hippocampal areas CA3 and CA1.* Science, 2004. **305**(5688): p. 1295-8.
8. Lee, I., et al., *Comparison of population coherence of place cells in hippocampal subfields CA1 and CA3.* Nature, 2004. **430**(6998): p. 456-9.
9. Vazdarjanova, A. and J.F. Guzowski, *Differences in hippocampal neuronal population responses to modifications of an environmental context: evidence for distinct, yet complementary, functions of CA3 and CA1 ensembles.* J Neurosci, 2004. **24**(29): p. 6489-96.
10. Leutgeb, J.K., et al., *Pattern separation in the dentate gyrus and CA3 of the hippocampus.* Science, 2007. **315**(5814): p. 961-6.
11. Madar, A.D., L.A. Ewell, and M.V. Jones, *Pattern separation of spiketrains in hippocampal neurons.* Sci Rep, 2019. **9**(1): p. 5282.
12. Knierim, J.J. and J.P. Neunuebel, *Tracking the flow of hippocampal computation: Pattern separation, pattern completion, and attractor dynamics.* Neurobiol Learn Mem, 2016. **129**: p. 38-49.
13. Bakker, A., et al., *Pattern separation in the human hippocampal CA3 and dentate gyrus.* Science, 2008. **319**(5870): p. 1640-2.
14. Baker, S., et al., *The Human Dentate Gyrus Plays a Necessary Role in Discriminating New Memories.* Curr Biol, 2016. **26**(19): p. 2629-2634.
15. Berron, D., et al., *Strong Evidence for Pattern Separation in Human Dentate Gyrus.* J Neurosci, 2016. **36**(29): p. 7569-79.
16. Yassa, M.A. and C.E. Stark, *Pattern separation in the hippocampus.* Trends Neurosci, 2011. **34**(10): p. 515-25.
17. Marmarelis, V., *Nonlinear dynamic modeling of physiological systems*. 2004: Wiley-IEEE Press.
18. He, F. and Y. Yang, *Nonlinear System Identification of Neural Systems from Neurophysiological Signals.* Neuroscience, 2021. **458**: p. 213-228.
19. Johnston, S.T., et al., *Paradox of pattern separation and adult neurogenesis: A dual role for new neurons balancing memory resolution and robustness.* Neurobiol Learn Mem, 2016. **129**: p. 60-8.
20. Wilson, I.A., et al., *Neurocognitive aging: prior memories hinder new hippocampal encoding.* Trends Neurosci, 2006. **29**(12): p. 662-70.
21. Guzowski, J.F., J.J. Knierim, and E.I. Moser, *Ensemble dynamics of hippocampal regions CA3 and CA1.* Neuron, 2004. **44**(4): p. 581-4.
22. Dimoka, A., et al., *Modeling the nonlinear properties of the in vitro hippocampal perforant path-dentate system using multielectrode array technology.* IEEE Trans Biomed Eng, 2008. **55**(2 Pt 1): p. 693-702.
23. Berger, T.W., et al., *Restoring lost cognitive function.* IEEE Eng Med Biol Mag, 2005. **24**(5): p. 30-44.
24. Gholmieh, G., et al., *An efficient method for studying short-term plasticity with random impulse train stimuli.* J Neurosci Methods, 2002. **121**(2): p. 111-27.

25. Berger, T.W., et al., *A hippocampal cognitive prosthesis: multi-input, multi-output nonlinear modeling and VLSI implementation.* IEEE Trans Neural Syst Rehabil Eng, 2012. **20**(2): p. 198-211.
26. Song, D., et al., *Nonlinear dynamic modeling of spike train transformations for hippocampal-cortical prostheses.* IEEE Trans Biomed Eng, 2007. **54**(6 Pt 1): p. 1053-66.
27. Song, D., et al., *Sparse Large-Scale Nonlinear Dynamical Modeling of Human Hippocampus for Memory Prostheses.* IEEE Trans Neural Syst Rehabil Eng, 2018. **26**(2): p. 272-280.
28. Cook, E.P., et al., *Dendrite-to-soma input/output function of continuous time-varying signals in hippocampal CA1 pyramidal neurons.* J Neurophysiol, 2007. **98**(5): p. 2943-55.
29. Koutsoumpa, A. and C. Papatheodoropoulos, *Short-term dynamics of input and output of CA1 network greatly differ between the dorsal and ventral rat hippocampus.* BMC Neurosci, 2019. **20**(1): p. 35.
30. Wang, W., et al., *Memory-Related Synaptic Plasticity Is Sexually Dimorphic in Rodent Hippocampus.* J Neurosci, 2018. **38**(37): p. 7935-7951.
31. Sclabassi, R.J., et al., *Nonlinear systems analysis of the hippocampal perforant path-dentate projection. I. Theoretical and interpretational considerations.* J Neurophysiol, 1988. **60**(3): p. 1066-76.
32. Berger, T.W., et al., *Nonlinear systems analysis of the hippocampal perforant path-dentate projection. II. Effects of random impulse train stimulation.* J Neurophysiol, 1988. **60**(3): p. 1077-94.
33. Berger, T.W., et al., *Nonlinear systems analysis of the hippocampal perforant path-dentate projection. III. Comparison of random train and paired impulse stimulation.* J Neurophysiol, 1988. **60**(3): p. 1095-109.

Chapter 7

An Electrical Engineer's journey to developing a vision for solving the neural code

7.1 Trajectory of an electrical engineer's investigation of the brain

Summarized in this dissertation are the key material of my PhD work, discussed in the order in which the work was done. While conducting these research works, in parallel, I was also gaining new insight through my electrical engineering coursework. The neuroscience research was initially primarily done through the lens of a neuroscientist. However, the engineering insight gained began to influence the work. Every chapter in this dissertation and my engineering coursework were critical to my trajectory of studying the brain and shaped the development of a longer term research vision.

My early work in neuroscience, primarily done through the lens of a neuroscientist, informed me of how neuroscientists think about the brain, state-of-the-art methodologies, experimental design, and the knowns and unknowns about the brain. Being integrated in the neuroscience environment and having acquired this information, I was able to identify avenues in neuroscience that could benefit from the electrical engineering insight gained. These avenues progressed from implementation of existing methodologies in system theory to the hypothesis driven questions in neuroscience, to adapting such methodologies, to finally, using the methodologies to study the brain comprehensively and inductively as a system.

7.2 Key take-homes from each chapter which contributed to shaping an overarching vision for solving the neural code

Chapter 2

In chapter 2, using behavior alone, we learned about the brain as a system, and how its specifications change with experience. We showed that mice running the obstacle course were

generally smarter than any other group. The study showed that the brain's prior experiences influence the encoding and/or storage of new experiences. From a system standpoint, prior processing and/or storage affects future processing and/or storage. The findings also pointed to a potential mechanism for this change in system specifications; numerous experiences may lead to efficient storage by organizing encoded information into a categorical structure making content readily accessible during memory recall, and new content readily integrated into this organized structure during learning new information.

These inferences about the system have implications for healthy and clinical populations. Positively challenging life experiences that are rewarded (obstacles in the rodent model) can make the brain smarter and protect it from the effects of aging and debilitating disease. In other words, positively challenging experiences that are rewarded matter, because they could lead to software and hardware changes that could help make wiser and more resilient agents.

Chapter 3

In addition to the discoveries discussed in chapter 3, three key lessons were gained about the brain system. First, like physical (non-biological) systems, the brain system response—neural activity measured during behavior—is composed of a transient and a steady state phase. Interestingly, these distinct phases are distinctly linked to behavior. It's the steady state period of the 20-40 Hz hippocampal response that predicted the accuracy of sequence memory performance.

Second, powerful analysis tools, such as machine learning, are critical to guiding investigations of high dimensional neural data. Machine learning and feature selection provided a means to comprehensively explore a high-dimensional feature space and identify useful neural information and its links to behavior. This undertaking may simply be impractical to do with the naked eye or through a series of univariate analyses. In our study, the *a priori* chosen

time windows of 20-40 Hz power, thought to most likely be linked to memory performance (right after the odor is presented, and right before a decision is made), showed no link. Had we not examined many features through machine learning feature selection, we would have missed the steady-state period that is critical for memory performance.

The last salient lesson from this study is the degree to which the 20-40 Hz signal is a general signal, nonspecific to a brain area. Based on the literature, 20-40 Hz is present in numerous brain areas and behavioral contexts (ex. motor, language, olfaction, auditory beat perception and more). We, however, encountered it in the hippocampus during a sequence memory task. This provides insight to the brain's coding scheme – it is possibly repurposing the 20-40 Hz signal and without interference. Finally, the seeming generality of this signal points to the need to understanding the basic principles of the brain–if we know the coding scheme of the brain, we'd know where 20-40 Hz fits operationally, and its behavioral consequences.

Chapter 4

Key take homes regarding neuroscience investigations were also obtained from the work discussed in chapter 4. When considering pattern separation (successful discrimination between memories), the field generally tended to focus on a few brain areas (hippocampal subfields). However, our work showed that numerous areas (hippocamps and neocortex) show distinct neural activity (increased 4-5 Hz power) during successful discrimination between memories. This pointed to the first lesson–the necessity of considering the remainder of the brain, since it is an interconnected network, when studying a given area. In other words, shying from readily ascribing key discoveries to the area in which they were discovered, and considering that the discovery may be generalizable to, or at least supported by, the rest of the brain.

In this chapter, information theory was used in a deductive approach, to estimate a direction of communication between two areas, and important discoveries were made to this

end. A bidirectional exchange between two key memory systems, the hippocampus and neocortex, is estimated by the phase signal of the 4-5 Hz frequency range. Although communication was largely bidirectional, small yet significant biases in the direction of the exchange varied in a behavior dependent manner (memory formation vs. retrieval and failed vs. successful mnemonic discrimination). It is possible that this 4-5 Hz signal is conducive to information transfer between distant brain areas since it is a slower frequency with a longer period that can account for the bottleneck of conduction velocity and multiple synaptic delays. An interesting extension to this work, however, would be an inductive use of information theory to identify the coding scheme of the brain. An undertaking would be to identify the sender/receiver maps across brain nodes and attempt to decode the coding scheme.

Chapter 5

In chapter 5, we adapted and applied two different techniques, phase-transfer entropy and Kalman filter-based granger causality, respectively, to estimate directional communication between the hippocampal and neocortical memory systems. The presence of these two approaches and their application on the same data is key for reproducibility – identifying findings that are robust to the approach. Future work is needed to test for reproducibility of findings made in chapter 4, by comparing the inferences obtained from each of these techniques.

More generally and inductively, considering the essence of the Kalman filter, that it updates the mean and covariance matrix to make predictions about new observations prompts a higher-level possibility. This possibility was previously presented in the literature –does the brain operate like a Kalman filter? Update its mean and covariance to predict the best next action?

Chapter 6

Finally, rather than conducting a study that prompts discussion of *likely* possibilities of brain operations or computations, we inductively and comprehensively captured these operations. Operations in the CA3:CA1 dendritic node were captured with limited a priori assumptions and in a manner non-specific to a behavioral state. We now know this is possible, to solve the code to one node of the brain.

We solved the code for one of the most heavily studied nodes. Despite this, we made a series of discoveries never before known. We found that contrary to popular belief, the CA3:CA1 node is nonlinear. Moreover, examining the system solution revealed how the system treats input spike patterns (attenuation and suppression based on spike timing); the system applies positive weights to spikes that are 2-5 msec apart that have occurred anywhere from 5-28 ms in the past relative to the current output. This speaks to the power of a comprehensive, agnostic and inductive approach – to ask the system to define itself, rather than the observer looking for what he/she assumes to be present in the system. The identified transfer function for the CA3:CA1 nodes generalizes across animals; the function can predict the entirety of the output waveform (instantaneous predictions) to arbitrary inputs delivered in a new set of animals with 93% mean external prediction accuracy.

Lastly, the node for which a neural code was solved serves as a model to many other nodes in the brain since its basic architecture is shared -- pyramidal cells which are all over the cortical telencephalon. Having done this successfully, in a model node, opens the door for doing this in the remaining brain nodes.

7.3 Possibilities that may arise from this work

The longer-term research plan is to apply system identification on additional nodes. For example, the pyramidal cell has two dendritic domains – apical and basal. We aim to estimate a transfer function for the basal domain and compare it to the identified apical domain transfer function. This will enable us to determine the differences in transfer dynamics at each of these

dendritic domains. Other inputs to the CA1 will be evaluated in the same fashion (e.g. temporoammonic pathway from the entorhinal cortex). Knowledge gained here will be integrated with the well-known physiology of synapses and dendrites such that a comprehensive understanding of the CA1's governing rules and dynamics can be established. In a similar fashion, this procedure can be repeated for the remaining nodes of the hippocamps (DG, CA3, subiculum), such that a complete model for the hippocampal formation can be achieved. A complete and accurate model of the hippocampal formation can be examined for emergent phenomena—the model can teach the observer about network capabilities and functions that arise from the network's basic design. Identifying the governing dynamics of the hippocampus is critical because it is a structure vulnerable to a host of different disorders.

The hippocampus is also a component in the larger circuit of the brain. We expect that the rules and dynamics of different nodes within this circuit to be different, depending on their functions and connectivity patterns. This level of complexity may complicate this endeavor, however, scaling upwards, transfer functions can be estimated for other circuit components and linked appropriately given knowledge of the anatomy. The end goal is to have this linking of nodes yield a single block diagram and its associated net transfer function. This single net transfer function would describe the governing dynamics for the brain. Scaling downward, to a single transfer function for a single synapse, we can identify the operations that transform input spike patterns to output spike patterns – the transfer function for the fundamental unit of communication of the brain.

By examining these transfer functions that explain brain governing dynamics, we will have more comprehensive models whose behavior could possibly enable us to build an adequate "theory of brain". In the same way, we can precisely define diseases and have a metric for reverting the obtained definition of disease to normal. For the first time, we can define the generality for the disease, and treat its biological basis, not its symptoms. The most important goal of the system theory-inspired full understanding of the human brain is to guide

individuals to having a good quality of life and to prevent and alleviate the suffering from debilitating neurological and neuropsychiatric diseases.