NEUROSCIENCE

NEUROSCIENCE FOREFRONT REVIEW

A. A. Fenton et al. / Neuroscience 529 (2023) 129–147

Do Place Cells Dream of Deceptive Moves in a Signaling Game?

André A. Fenton, [a,b]* **José R. Hurtado,** [a] **Jantine A. C. Broek,** [c] **EunHye Park** [a] **and Bud Mishra** [c,d,e]

[a] *Neurobiology of Cognition Laboratory, Center for Neural Science, New York University, New York, NY, USA*

[b] *Neuroscience Institute at the NYU Langone Medical Center, New York, NY, USA*

[c] *Departments of Computer Science and Mathematics, Courant Institute of Mathematical Sciences, New York University, New York, NY, USA*

[d] *Department of Cell Biology, NYU Langone Medical Center, New York, NY, USA*

[e] *Simon Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA*

Abstract—We consider the possibility of applying game theory to analysis and modeling of neurobiological systems. Specifically, the basic properties and features of information asymmetric signaling games are considered and discussed as having potential to explain diverse neurobiological phenomena; we focus on neuronal action potential discharge that can represent cognitive variables in memory and purposeful behavior. We begin by arguing that there is a pressing need for conceptual frameworks that can permit analysis and integration of information and explanations across many scales of biological function including gene regulation, molecular and biochemical signaling, cellular and metabolic function, neuronal population, and systems level organization to generate plausible hypotheses across these scales. Developing such integrative frameworks is crucial if we are to understand cognitive functions like learning, memory, and perception. The present work focuses on systems neuroscience organized around the connected brain regions of the entorhinal cortex and hippocampus. These areas are intensely studied in rodent subjects as model neuronal systems that undergo activity-dependent synaptic plasticity to form neuronal circuits and represent memories and spatial knowledge used for purposeful navigation. Examples of cognition-related spatial information in the observed neuronal discharge of hippocampal place cell populations and medial entorhinal head-direction cell populations are used to illustrate possible challenges to information maximization concepts. It may be natural to explain these observations using the ideas and features of information asymmetric signaling games.

Key words: game theory, information Asymmetry, hippocampus, entorhinal cortex, spatial cognition, navigation, cell assemblies.

INTRODUCTION

A child might ask ‘‘how does the brain work?” Because computers are now second nature and the brain clearly computes, one might be tempted to draw on analogies between brains and computers to answer the child, but an agile self-respecting neuroscientist is more likely to answer, ‘‘well, you know, the brain is the most complex system we know of, so it should not surprise you that we do not understand how it works.” Dinner party guests often ask ‘‘how does memory work?” to which our neuroscientist might answer, ‘‘your neurons make a molecule named PKMzeta that is crucial,” or ”synapses change their effectiveness to store memories,” or ‘‘when you experience something, the neurons in your brain discharge electrical signals in specific patterns, those patterns replay when you recall the memory,” or ‘‘there is a part of the brain called the hippocampus, it is where you store memories, until they are transferred to the neocortex.” Depending on the quality of the wine, the neuroscientist might add ‘‘it doesn't seem at all like computer memory.” While each of these proximate explanations is in a sense standard, it is remarkable that each involves a distinct level of biological organization, where a great amount of self-consistent, rigorously-obtained detail is known within the level, but rather little is known about how to connect the phenomena between the levels and advance an ultimate explanation. Such an explanation would also encompass evolution, development, and genetics. Indeed, concepts like transcription, translation, and post-translation modifications like phosphorylation that operate at the nanoscale and minutes-long timescales of genes and macromolecules, may appear off the mark when

*Corresponding author.

E-mail address: afenton@nyu.edu (A. A. Fenton).

Abbreviations: PKMζ, protein kinase M zeta; CA1, Cornu ammonis area 1; CA3, Cornu ammonis area 3; DG, dentate gyrus; STDP, spike timing dependent plasticity; E, excitatory; I, inhibitory; Infomax, information maximization; DS_M, MEC-originating dentate spike; LIF, leaky integrate-and-fire; DREADD, designer receptor exclusively activated by designer drugs; SG, slow gamma; MG, mid-frequency gamma; SGdom, Slow-gamma dominance.

https://doi.org/10.1016/j.neuroscience.2023.08.012
0306-4522/© 2023 The Author(s). Published by Elsevier Ltd on behalf of IBRO.

explanations turn to the dynamics of electrical discharge through neuronal circuits and their computations.

The experimentalists amongst us work to understand how the hippocampus-entorhinal cortex neuronal circuitry operates. We identified that the persistent, autophosphorylating enzyme, protein kinase M zeta (PKMζ) is translated from mRNA at postsynaptic dendritic sites that were activated during recent learning and that this metabolic synthesis is crucial for long-term memories to persist (Sacktor, 2011; Tsokas et al., 2016). We identified that increased PKMζ expression and synaptic strengthening persists at a subset of hippocampal-entorhinal synapses, for at least a month, so long as the memory persists (Pavlowsky et al., 2017; Hsieh et al., 2021). We also validated the hypothesis that memories of a place in which discomfort was experienced are recollected when slow gamma oscillations originating at the Schaffer collateral synapses of hippocampus subfield CA1 dominate mid-frequency gamma oscillations that originate at the *stratum lacunosum moleculare*. This competition manifests as memory-associated neuronal ensemble ''place cell'' discharge that resembles the ensemble location-specific discharge at the recollected place, despite the subject initiating the recollection from the current location, which is a physically different place (Dvorak et al., 2018; Dvorak et al., 2021). Although we study these memory phenomena at distinct levels of biology, in the same laboratory, we still shy away from designing experiments to test hypotheses across the different levels of biology. This is in large part because we lack conceptual frameworks that are useful for describing, organizing, or understanding the cross-level neurobiological memory phenomena (see Bell, 2008).

THE CHALLENGE OF CROSS-SCALE ANALYSIS AND UNDERSTANDING

Brains are self-organizing and changing. Simply using a brain changes it structurally as well as functionally, so that it will operate differently in the future, as we recently showed in mice (Chung et al., 2021). We also have incomplete knowledge of the myriad neurobiological processes that result in neuronal implementations, algorithms, and even the computations involved in the formation and maintenance of memory, three features which contribute to our overall understanding of cognitive processes (Marr, 1971). Even in artificial computing systems where we fully know the implementation, algorithms and computations, neuroscience techniques commonly used to infer functions from recorded data are ill-equipped for cross-scale analysis in the relatively simple architecture of a microchip, let alone a human brain (Jonas and Kording, 2017). We believe this lack of integration in the study of memory severely limits the generalizability of conceptual frameworks across scales.

Moreover, although neuronal datasets have become larger and more comprehensive in the past decade, the unavoidable under-sampling of neurons has motivated an interest in a theoretical framework that can guide predictive models from sparse data with the intention to unify observations at many scales of analysis (Bassett and Sporns, 2017). It is here where a new conceptual framework can have substantial utility if it explains how a unique biochemical phenomenon at one scale is constrained by and enabled by distinct phenomena across scales. Such a framework should have predictive power beyond scale-specific and substrate-specific processes, similar to universality classes (Ódor, 2004) that illustrate how unrelated materials like a flock of birds and a network of neurons behave in a statistically similar manner near critical conditions (Fruchart et al., 2021).

Considering the need for a framework that encompasses the interactions between emergent neurobiological phenomena and their constituent elements, we believe the concepts and analytical framework of signaling games are well suited for this cross-scale analysis, primarily due to its dynamical and decentralized modeling approach. The framework has been applied to diverse domains of enquiry, where information is asymmetrically distributed among freely interacting entities. Applications have included the evolution of the genetic code (Jee et al., 2013), macromolecular signaling cascades and immunology, economics (Spence, 1973) and financial systems, and cyber security and internet governance (Casey et al., 2019), which in our opinion bodes well that the framework may fit the bill to meet the needs of our cross-scale challenge to understand the neurobiology of memory in particular, and maybe even brain function in general (Rosser, 2003).

In the original (non-cooperative) formulation of game theory (Nash, 1950), the information was considered symmetric between players, in that both players reveal exactly the same information, and use that information in simultaneous strategic choices. Accordingly, in an information symmetric system, information is revealed faithfully, but in more general signaling systems modeled as signaling games, signals can be enhanced with additional or more reliable information. The signal can also be disinformation — deceptive and distorted with respect to the receiver's expectations. Fig. 1 presents the extensive form schematic of an information asymmetric game to illustrate how a signaling game can operate. We do not intend to model anything here, much less a neural system with many elements and complexity in the systems operations. The scores in this illustrative example are arbitrary, set by the circumstances of the game, in this case by the authors. Player I (circle PI) is the Sender and Player II (circle PII) is the Receiver. Consider each to be a pair of successive neuronal processing nodes. The schematic begins with the center vertical line, where nature (N) selects the sender type on each instance of the game. Each bifurcation on the diagram indicates Player I's (Left vs Right) or Player II's (Up vs Down) choice, after which there will be a payout for each player that depends on their respective moves and utility functions. Player I can be one of two types in the instance of each game, i.e., a transmitter of sensory representations (red) or of memory representations (blue). Imagine Player II is involved in the motor output in response to an activity signal from Player I, but Player II does not know if Player I's signal is a sensory or a memory signal type. Upon receiv-

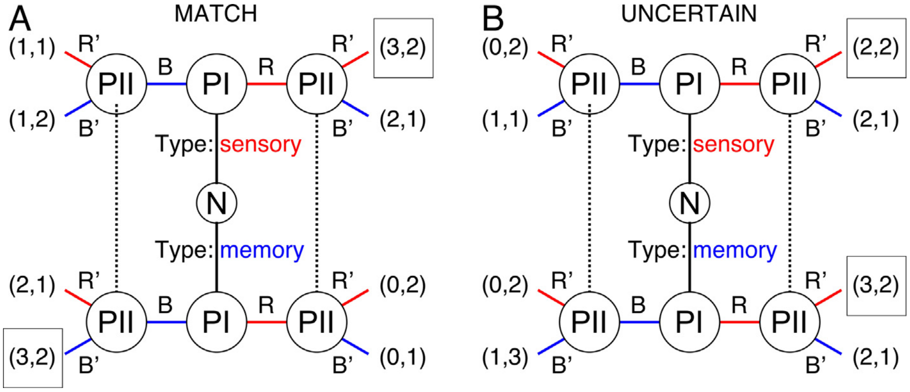

Fig. 1. Two example information asymmetric signaling games to illustrate the game's operation and ability to do information processing based on the payout structure. (**A**) The MATCH game incentivizes 'honest' signaling by Player I. Player II is incentivized to match Player I's type with an identifying response because Player I reveals its type faithfully to Player II. When Player I is type sensory it is likely to signal R (Sensory-R payouts > 1) instead of B (Memory Sensory-B payouts = 1) in hopes that Player II will guess R', which Player II is more likely to do because the (R,R') payout = 2 instead of B' in which the (R,B') payout = 1. When Player I is type memory it is likely to signal B (Memory-B payout > 0) instead of R (Memory-R payout = 0) in hopes that Player II will guess B', which Player II is more likely to do because the (B,B') payout > 1 instead of R' in which the (B,R') payout = 1. Player II correctly reports Player I's type by generating a corresponding signal that becomes reliably correlated with the sender's type. There are no profitable deviations for Player I given Player II's strategy. This corresponds to a so-called separating (Nash) equilibrium because the sensory and memory nodes have adopted distinctive signals (marked by rectangles), allowing the Player II node to distinguish them. (**B**) The UNCERTAIN game has a different payout structure that incentivizes Player II to match Player 1′s sensory responses but not memory responses, resulting in suboptimal conditions for Player II that decorrelates the Player I signal from the sender type. When Player I is type sensory, it is likely to signal R (Sensory-R payout > 1). Player II's guess should be R' because the (R,R') payout = 2 instead of guessing B' which has (R,B') payout = 1. However, when Player I is type memory it is now also incentivized to signal R (Memory-R payout > 1) instead of signaling B (Memory-B payout < 2). Player II is therefore compelled to signal R' because the (R,R') payout is greater than the (R,B') payout, even though Player II could have done better if Player I revealed a distinguishing B signal since the (B,B') payoff = 3 for Player II. However, given these payouts, Player I would be worse off (Memory-B payout < 2). The fact that Player I benefits from ambiguating signals at the expense of the receiver, Player II, is the hallmark of deception. The B and R signals are now decorrelated from the sender type in proportion to the number of memory types in the population, breaking the separating equilibrium. Given that Player II does not know Player I's type, Player II will reliably report R' in response to Player I's signals, whether Player I's activity originates from the sensory or memory node. This corresponds to a so-called pooling (Nash) equilibrium because the sensory and memory types have adopted the same signals, preventing the Player II node from distinguishing them, and pooling the two types of senders into the same response strategy (marked by rectangles).

ing the signal, Player II responds by sending either motor signal R' or B'. What signal each player is likely to send depends on the individual agent's utility payouts, which is a function of the action of both players. Two possible signaling games are depicted in panels A and B. The possible interactions that define each game are identical, but the two games differ in the two payout structures, represented as coordinates in ordered pair form: (Player I, Player II). One player must respond strategically to the other's signal, which can keep unrevealed information private. In game theory, the correlations between sender signals and receiver responses correspond to priors regarding what type of sender is affiliated with a given signal, and since interests are not necessarily aligned between the information-rich senders and information-poor receivers, these priors are effectively "beliefs" that determine the strategies for signaling behavior and may be updated after each subsequent interaction. What generates different signaling games is the information processing that occurs enroute to a particular receiver. In information asymmetric signaling games the sender may know the meaning of the signal, but the receiver is uncertain and possibly completely ignorant about the meaning (Smith, 2000; Rosser, 2003; Dongen, 2006; Jee et al., 2013). These additional considerations introduce information asymmetry between players, and this asymmetry facilitates more complex interactions between agents.

Note that a receiver in one communication can then act as a sender to transmit signals further through the system in a chain or network-like fashion at each stage, with direct or indirect feedback coupling. As we will discuss presently, this situation represents precisely the signalling condition of the neurons that constitute a nervous system, particularly in association areas like the hippocampus where drifting and multi-stable activity is a feature of its memory and learning function (Kelemen and Fenton, 2013; Sheintuch et al., 2020; Chung et al., 2021). Although popular neural network models of brain computations rely on circuit connections that explicitly separate sender types by the information that they relay, we present a network model demonstrating how parameters of individual neurons (Excitability, Inhibitory Gain, Time Constants etc.) can be collectively tuned to generate multiple structured representations from otherwise unstructured connection architectures (Pehlevan and Sompolinsky, 2014). With regards to signaling games, our model demonstrates the existence of multiple informative states that emerge as stable responses to a continuous variable (position) signaled by random inputs. Any different set of random inputs generates states that are representationally equivalent, in that they model the desired variable, but would not be equivalent in biological terms with respect to the metabolic and signaling costs incurred by individual neurons. For instance, it is clear from Fig. 2 that a subset of neurons fail to fire reliably in the random input-weight network, which is likely a substantial cost for neurons (Laughlin et al., 1998; Chintaluri and Vogels, 2022). This demonstrates the potential utility of a signaling game framework, because informative cell assemblies arise due to individual signaling parameters in the absence of labeled lines, resulting in multiple solutions to information processing and an opportunity to compete for network resources across the multiple informative equilibria. Due

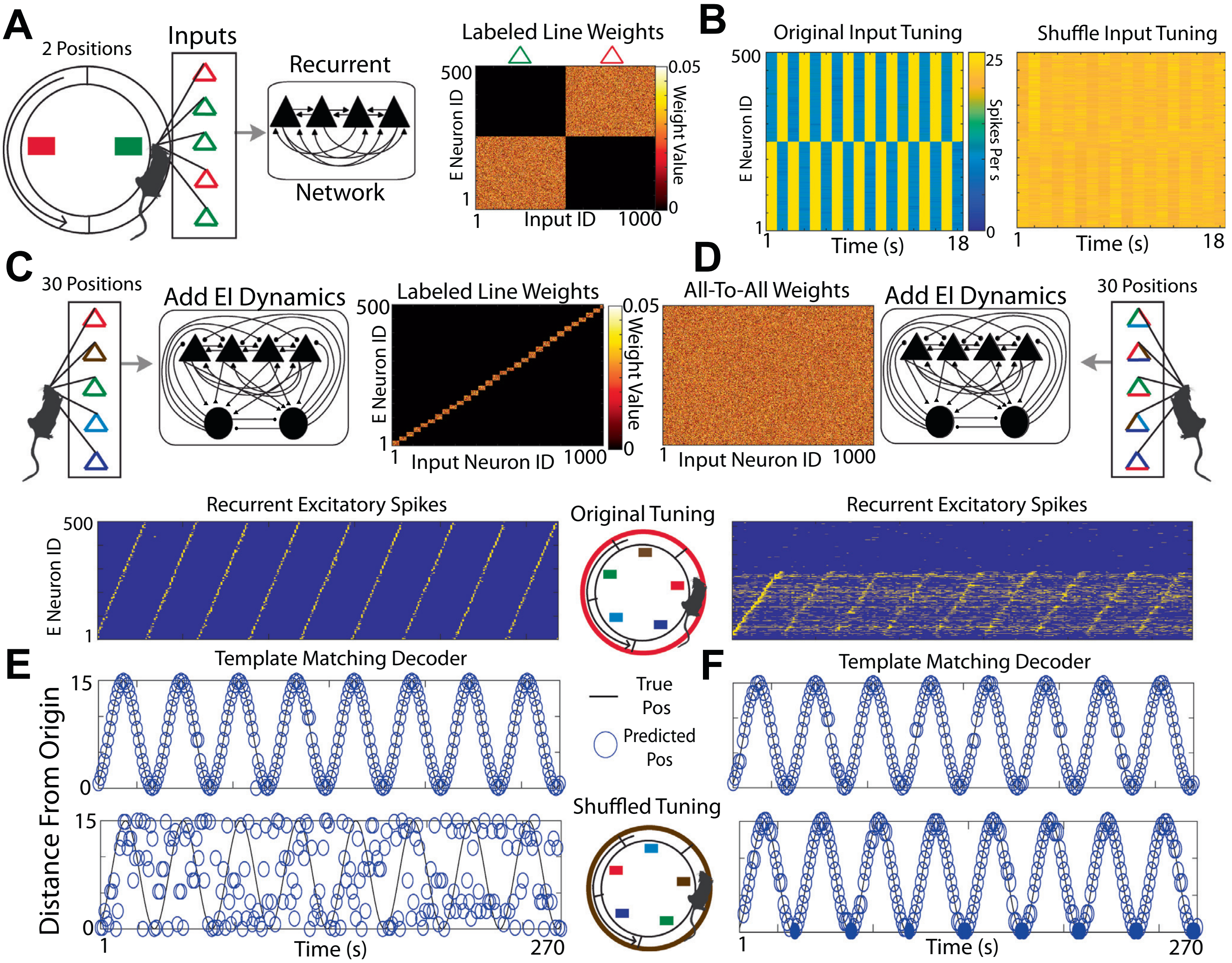

Fig. 2. Pattern formation in randomized leaky integrate-and-fire (LIF) network with constant recurrent connection weights. This illustrates the emergence of informative signaling equilibria from noisy inputs with mixed selectivity and random, all-to-all inputs. (**A**) Simplified LIF model of simulated positional inputs ($P = 2$) around a circular track, differentially encoding the positions. These inputs provide labeled-line feedforward positional signals to the recurrent excitatory (E) network. (**B**) The E firing rates reflect the enforced positional tuning of the input weights (left), however, the network is not capable of differentially encoding the positions when the input tuning is randomized (right). (**C**) Top: Elaborated $P = 30$ position LIF Recurrent Network model with inhibitory (I) all-to-all connections within the recurrent network, and its corresponding labeled line connection structure. Bottom: Recurrent E firing rates encode the position trajectories, and representations are made sparse by inhibition. (**D**) Top: Randomized LIF Recurrent Network model as in panel C with random input weights (left) and feedforward inputs tuned to multiple (20%) locations. Bottom: Recurrent E firing rates encode the position trajectories. E neurons are ordered according to within-position spike primacy, where the first units to spike at a given position are ordered before the first units to spike at the next position. Only the first 16 neurons to spike at a given position were considered in the ordering. Note that this cell-specific selectivity occurs despite each E neuron receiving signals from every input neuron with randomized connection weights. Template-matching decoding accurately predicts the position trajectories from recurrent E activity in (**E**) the labeled-line and (**F**) the random-input recurrent networks. Top: The position-averaged activity from separate simulations using the same inputs and architecture serves as the template for the decoder. Bottom: When positional tuning is randomly reassigned in the inputs, the template matching decoder fails to decode the new trajectories from the recurrent E activity and template in the labeled-line recurrent network but succeeds for the random-input network, which generates new network firing activity patterns that correspond to the new separating equilibria of the network with distinct positional tuning across the recurrent E cells.

to the differential allocation of signaling costs, signaling strategies determine the winners and losers of the signaling game, as well as the overall persistence of the signaling game. Separate work will model changing strategies in this network, in this perspective we focus on implications of the network model and the signaling game framework generally.

SIGNALING GAMES

We previously used a recurrent network with excitatory (E) and inhibitory (I) leaky-integrate-and-fire units to demonstrate that random input patterns sampled from locations along a ring can be represented by position-tuned responses only if spike timing dependent plasticity

(STDP) learning rules are activated amongst the E-I and I-E connections (Levy et al., 2023). Because STDP at the E-E connections was not effective, it is unclear whether the network learned position with the connection strengths, or alternatively, if the learned connections enabled the network to process the place information without storing it (O'Reilly et al., 2019). Here we modified the model to illustrate the value of the signaling game framework by providing a network example that cannot be easily understood from the standard view that the particular set of connection strengths between neurons in a network store information and define the information channels that these neurons encode (Zhenrui et al., 2022; Levy et al., 2023).

We begin with so-called labeled-line architectures as described in sensory systems and the findings of topographic maps, which have strongly influenced the field's thinking about the efficiency of hard-coded channels for encoding and decoding the maximum amount of information available (Fig. 2(**A–C**)). If each network unit only receives strong input from a particular stimulus subset it is then straightforward to understand how those network units follow the input activity patterns to the recurrently-connected units. However, as information becomes more integrated across senses and memory, it is less clear how information channels could be structured topographically or systematically in association regions like the hippocampus (Redish et al., 2001; Levy et al., 2023). Multi-modal information converges from both cortical and subcortical regions in hippocampus, with no obvious way to combine or generate labeled-lines or other similar functional topographies. By definition, labeled lines are inflexible and thus maladaptive to a system that must remap its patterns of cofiring activity to generate distinct memories and new experiences (Fig. 2(**B,E**)). We organized the recurrent network with random all-to-all feedforward input connections, without any connection strength plasticity. The network can nonetheless generate selective unitary responses that represent each unit's strongest inputs within the temporal structure of the input activity (Fig. 2(**D**)). The momentary input can be readily decoded as a unique pattern of network activity (Fig. 2(**E**)), even with the random-input network (Fig. 2(**F**)). This result is surprising because, instead of using a learning rule to set the recurrent weights, some high and some low, we set all the E-E (0.20), I-E (0.05), and E-I (0.05) weights to constants. Moreover, each E unit receives simultaneous signals from all input units, each with random weights, and yet the E units reliably settle into configurations of stimulus-tuned assemblies so long as the network is EI balanced. This effect occurs in part because recurrent inhibition prevents global runaway excitation, allowing only a subset of neurons to be active for any given stimulus pattern. Consequently, the network establishes a reliable configuration of positional tuning despite having no positionally tuned input connections and despite having no variety in the recurrent connection weights. Unlike the labeled-line network configuration (Fig. 2(**E**)), the random-input network even makes the network robust to remapping of the feedforward input tuning (Fig. 2(**F**)). This situation holds because the random assortment of positional signals onto every unit allows for a diverse array of new stimulus-dependent configurations.

We do not assert this random-input network configuration is realistic, rather we use it to illustrate that information representations can emerge dynamically and agnostic to explicit information tuning in either the inputs or the connection weights, which is a challenge for standard notions. The Fig. 2 example includes a predetermined EI balance for the network to settle into stable configurations. In fact, by allowing STDP rules to self-organize the EI balance in the network through changes in excitatory and inhibitory interactions, the representations become robust to arbitrary initializations of the EI balance (data not shown). This observation suggests that the organization of E cells, into informative stimulus-tuned assemblies, may be a dynamical after-effect of internal recurrent dynamics that self-organize based on local spike-time differences that are only circumstantially coupled to information about the external world from where inputs originate. These phenomena can be analytically interpreted from the perspective of a non-reciprocal dynamical systems theory (Fruchart et al., 2021) whereby the interactions between non-reciprocal E and I units are well-suited for pattern formation.

This example of the random-input network that challenges standard intuitions has motivated us to consider alternative frameworks, which is why we hypothesize that signaling game theory can offer a powerful complement for understanding how signaling agents can settle into informative coactivity patterns in the absence of global optimization functions. The adaptive utility of relaying a particular signal, such as the local minimization of spike-time differences within assemblies via STDP, determines the structure and function of any signaling game. In a signaling game, the collection of senders and receivers interact strategically via signals, each trying to maximize local functions that are dependent on the actions of other signalers. These functions are local utilities that determine the ability of the signaling agent to persist in the signaling game. Those utilities are influenced by the current signals in the game, as well as the subjective probabilities about the implication of the transmitted signals. These subjective probabilities are only abstractions representing how an agent will interpret a given signal based on the past action histories. Note, it is important to recognize that these subjective probabilities do not require an epistemic agent. The strength of a synaptic connection is in essence a subjective probability measure of the signaling value that is transmitted by the presynaptic neuron to the postsynaptic neuron, and this subjective probability may not reliably communicate the true information value of the signaling therein. What matters instead is that the postsynaptic cell is biophysically compelled to expend its metabolic resources by responding with a postsynaptic change in ion flux, which is measured as a membrane voltage deviation before returning to the resting potential. With additional metabolic expenditures, that synaptic connection can be persistently strengthened or weakened, adjusting the metabolic cost of receiving subsequent patterned inputs (Tian et al., 2008). In fact, non-epistemic signaling games can also occur in cases without any clear adaptive advantage, such as the signaling game between cancer cells and the immune system combatting the metastatic signaling cascades. Here, both the cancer cells and the immune cells have signaling utilities that govern local interactions, and these utilities are dynamically interacting in a manner that is only diffusely governed by natural selection (Casey et al., 2021). Within an organism, it is instead the signaling stable strategies, akin to the ''evolutionary stable strategies'' proposed by John Maynard Smith, which permit certain signaling games to persist over time, often at deleterious costs for the organism (Smith, 2000). Although the specific instantiations of signaling systems differ, there are general properties of signaling systems that are universal across scales. Fundamentally, the activity of any signaling system is defined by its signaling conventions, which is attractive for a cross-scale framework to formalize neuroscientific understanding.

Notice that, in an important way, establishing a convention is different from transmitting information, akin to the distinction between memory for information storage and memory for information processing (O'Reilly et al., 2019; Chung et al., 2021). In fact, establishing a convention is a prerequisite for information transmission. We highlight this distinction because a standard way of conceptualizing neuronal systems, and the brain, is as an information maximization organ. Indeed, popular neural network models believed to be relevant to brain function often impose information maximizing constraints onto the neuronal architecture and connection updates (Linsker, 1997). These top-down constraints are typically objective functions which change connection strengths in order to minimize the error in the output readouts with respect to a ground truth, which the system cannot in principle know. Even higher order models of cognitive processes, like attention, often use global optimization functions explicitly defined to minimize network prediction error, and prediction error variance, for a predetermined type of input (Andersen, 2022). However, little is known about how normative learning functions could emerge from biochemical agents to adequately realize relevant representations globally in a self-organizing network that initially lacks these normative constraints, highlighting the challenge of extrapolating from one level of biology to another. Other common network models utilize global dynamical equations of neural activity to constrain the patterns of firing to ''attractor states.'' These attractor states can be fixed points, lines, or manifolds which constrain the space of observed neuronal co-firing patterns in the network. In these ''attractor networks,'' the constrained space of activity patterns can function as reliable coordinates of neuronal representations, or even mechanisms of memory encoding. Although these attractor networks generate multiple stable states of activity, these path constraints are typically imposed top-down on the network architecture, and little is known about how a biological brain can self-organize into reliable attractor states from an initially unconstrained network configuration. Furthermore, some popular examples of attractor states being generated in networks through learning, such as the Hopfield network models, are forced to rely on global optimization of the Hopfield energy as a computation to be minimized across all neurons in the network, in contrast to the individualized maximization of utilities evoked by the signaling game perspective (Hopfield and Tank, 1985; Samsonovich and McNaughton, 1997).

This paper explores the merits of conceptualizing neuronal systems as signaling organs engaged in signaling games where conventions emerge spontaneously. These conventions that are analogous to attractor states can sometimes play roles that are akin to objective functions that constrain activity of the signaling system towards a representation of an external variable. The premise is there is no ''true'' global objective function for the network, nor a predetermined path constraint on activity. Each individual signaling agent simply behaves according to its individual signaling strategies, and changes those strategies based on individual utilities like metabolic costs (Laughlin et al., 1998; Chintaluri and Vogels, 2022). Signaling conventions are emergent phenomena and have varying degrees of reliability, where faithful information processing within a given convention is only one of multiple possible states of the signaling system.

The utility functions of individual signaling agents can determine system's collective behavior

A tractable definition of utility functions determining the behavior of individual signaling agents will facilitate the quantitative modeling of signaling strategies, and in the absence of a pre-specified utility function, its resulting maximization strategies can be plausibly inferred from past action histories and an enumeration of the possible states of the agent (DeDeo et al., 2010).

Given each agent in the game has a utility function that it will aim to optimize through signaling, the system of agents is likely to settle into one of several homeostatic states, which in game theory correspond to Nash equilibrium states (Binmore, 2007). A Nash equilibrium is a profile of strategies such that each player's strategy is an optimal response to the other players' strategies. An equilibrium state is explicitly defined as one in which each agent's behavior is an optimal response with respect to their utility function such that on the whole, no agent can gain more utility by changing their behavior.

It is important to highlight that Nash equilibria correspond to signaling conventions that are adopted by the signaling system, and that these equilibria may or may not promote optimal information transmission across the system. Signaling conventions operate like how social conventions govern the interactions between strangers that meet. To make this essential distinction clear in the context of information processing, consider the notion of an object in a computer program, or a piece of software like a web browser. In each case the software establishes a set of conventions as regards data types and even what operations can be performed on the data types. As in a signaling game, those conventions do not specify or define what specific information is represented or signaled to the receiving piece of code or the user. As in the signaling game in Fig. 1, and as we hypothesize in neuronal systems, those conventions allow for information processing to occur, and constrains what and how information can be processed, because of the conventions, although the conventions do not specify the information itself, that is another matter.

Batesian mimicry in snakes illustrates a biological game theoretic example in which conventionally informative signals, meant to signal ‘‘possession of dangerous venom,” can be co-opted by snakes that possess the conventionally informative color patterns without the metabolically costly venom that reinforces the convention (Casey et al., 2020). The authors characterize the mimics as ‘‘deceptive” signaling agents that obtain utility at the expense of the reliable convention. However, the utility is not absolute; it depends on the prevalence of deceptive agents in the population. A high prevalence of non-venomous types (deceivers) within a population (senders) with conventionally venomous-signaling patterns allow predators (receivers) to adaptively reduce the significance of these color patterns and they consequently increase predation of the snake population. Because increasing prevalence of deception reduces the salience of the signaling convention, and the cost of increased predation is dear, the tradeoff sets an equilibrium limit on the proportion of deceptive agents in a population that is operating under unchanging and stable conventions.

Honest, costly, and deceptive game theoretic signaling may seem intuitive in the domain of animal evolution, nonetheless, these principles are broadly generalizable, and we argue, they are translatable to the domain of neurobiology. Most neuroscience studies involving signaling games address cognition, honesty, and deception at the level of human social interactions (Jenkins et al., 2016), but these concepts of signaling games have also been applied to analyze biological phenomena at the non-behavioral sub-cellular scale of genes, RNA, and proteins (Massey and Mishra, 2018). Our fundamental conjecture, the central hypothesis of our program is that once the persistence of a signaling convention confers reliable adaptive utility to an organism, then there will be selective pressure against destabilizing the convention. Moreover, although evolutionary fitness determines selective pressure at the organismal level, agent-based interactions between genes, proteins, or neurons allow for complex and dynamic behavior within a signaling system. These interactions, at the level of the signaling agents, could appear competitive or cooperative in terms of their local utility functions. Nonetheless, it is the joint global effect on organismal fitness that allows these signaling conventions to persist through evolutionary timescales. Therefore, selection can act on parameters that regulate the degree of cooperation and competition between signaling agents in a single signaling system, as well as interactions between signaling systems in a single organism. Furthermore, the degree of competition or cooperation could be dynamic, and sensitive to environmental changes within a lifetime, with the flexibility of this dynamic under the regulatory influence of natural selection. Thus, if it is true that groups of organisms can alter their collective behavior through reinforcement and variation of communicable signals, we surmise it is also true at all levels of biological organization, including the levels of genes and biochemistry, cell biology, development, neuronal circuits, anatomically- and functionally-defined neuronal systems. While the particular signaling games, their local utility/cost functions and conventions may differ for each level of analysis, the analytical framework is universal and compatible within modern evolutionary theory.

We believe the ambition of a universal analytical framework to be an essential and noble goal because, as we indicated at the start, it remains poorly understood how local biochemical interactions give rise to self-organizing networks capable of sophisticated computational and cognitive capabilities. This is in part because hypotheses regarding brain computations are filtered through assumptions about the cognitive function of those computations, whereas decentralized signaling games with the proper structure generate emergent computations in nontrivial ways solely through local, agent-based interactions. We are heartened because, in simple computational models programmed as signaling games involving categorical discrimination of visual inputs, information used by agents for categorical discrimination of images can be unrelated to the conceptual features that researchers use to define the category (Bouchacourt and Baroni, 2018). Put another way, if naive signaling agents can find mutually beneficial conventions to communicate distinctions, then they can interact locally to optimize their utilities regardless of the nature of those distinctions, thus generating correlations between signals and responses which lead to informative population activity.

This process of generating conventions through distinctions can reinforce (or destabilize) the existing conventions and have consequences which feed back to the local signaling agent’s toolkit of utility/cost functions. In this manner, the reliability of a convention for transmitting faithful distinctions is what differentiates the types of Nash equilibria found in any given signaling game. One type of Nash equilibrium is called ‘‘babbling,” during which the agents interact via information-poor signals. Such an equilibrium, while not effectively purposeful, allows the system to explore possibilities. By chance, some of these possibilities allow the investments of costly signaling to improve the reliability of signaling, allowing receivers to predict sender type more reliably, raising the utility of some agents, and changing the type of equilibrium to what is called a ‘‘separating” equilibrium. As shown in Fig. 1(**A**), a separating equilibrium is conceptually similar to a discrimination as accomplished by a competitive neuronal network (Fig. 3), whereas a pooling equilibrium accomplishes the opposite because the different sender types elicit the same response as described in Fig. 1.

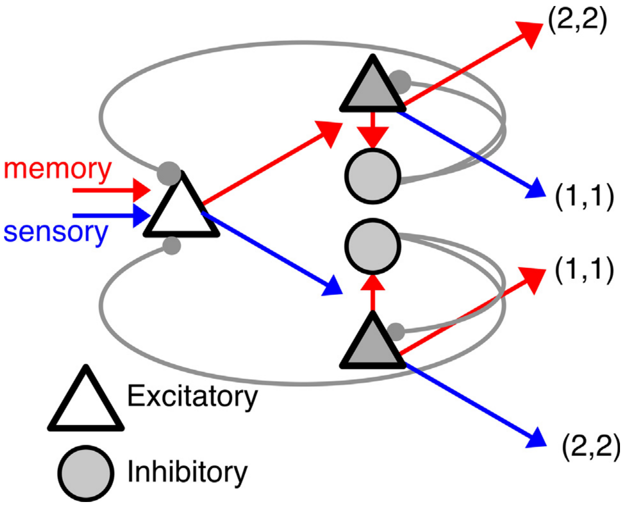

Fig. 3. Schematic neural network, where each neural element is a processing node (triangle, circle) and can represent a single neuron or an ensemble of neurons. Given its recent inputs player 1 (open triangle) can generate a representation of type sensory or memory and signal either R or B to the next processing node, the "receiver." The receiver node is illustrated with feedback and feedforward inhibition via tunable inhibitory synapse-like connections, and it will itself signal either R' or B', according to the utilities of the individual processing nodes. The utilities of the two nodes have been set in this example such that the utility of the R and B responses are equal. Consequently, the network will not settle into a stable pattern of signals unless a bias is introduced in the inputs or the connection strengths. The rationale for setting the utility functions for the neuronal network nodes is crucial to the implementation of the game theoretic approach that we are contemplating.

What Marr defined as "pattern separation," corresponds to a separating equilibrium, recognized as distinctively informative patterns of cofiring in neuronal networks (Marr, 1971). What Marr defined as "pattern completion" corresponds to a pooling equilibrium, observed when cofiring patterns are similar across different occasions. Accordingly, neuronal population discharge can organize into multistable states of separating, pooling, and babbling equilibria that can be distinguished by cofiring relationships and their differences (Schneidman et al., 2006; Park et al., 2019; Levy et al., 2023). By definition, when patterns of neuronal cofiring are far from separating equilibria the signaling convention is more readily adapted to changing conditions without either the need or cost of adopting a new signaling convention, which may have advantages for memory encoding that have not been recognized by standard concepts including attractor dynamics, which have tended to emphasize separating equilibria despite sometimes contrary observations (Knierim, 2002; Lever et al., 2002; Guzowski et al., 2004; Leutgeb et al., 2004; Wills et al., 2005). In fact, recent large population recordings from hippocampus have revealed that cofiring patterns in the neuronal population are largely indifferent to distinct environments, even though environment-specific information can be decoded from the neuronal population (Nagelhus et al., 2023; Levy et al., 2023). Such patterns of activity may be relatively straightforward to explain in the game theoretic framework, but they challenge standard notions of neuronal information representation.

Separating equilibria are characterized by stable conventions wherein strategically optimal signaling is distinct for distinct information types (Crawford and Sobel, 1982). In essence, once the agents establish verifiable signals with net utility outweighing the cost of such signaling, then the system will settle into separating equilibria (Sobel, 2007). This would be the case for an organism that possesses neurons that self-organize into reliable cofiring assemblies, likely at the substantial expense of "costly signaling," as we will elaborate presently. In terms of neuronal activity dynamics, it is easy to describe the activity pattern of a competitive network as a separating equilibrium when a subset of cells is vigorously coactive and a competing set of cells have transitioned to being relatively inactive as is observed in attentional and other functional descriptions of cortical networks (de Almeida et al., 2009). Fig. 3 illustrates such a competitive network, which by balancing excitation and inhibition has a strong tendency to adopt activity patterns that correspond to separating equilibria (Kaski and Kohonen, 1994).

THE INFLUENCE OF INFORMATION ASYMMETRY IN SIGNALING GAMES

Information asymmetric signaling game theory assumes no central or overall governance mechanisms. Instead, the rules by which the agents behave apply locally and govern local interactions between scale appropriate signaling agents. These conventions are the kinds of local rules that explain much of the behavior and dynamical structures of starling murmurations that can be accounted for by each starling following three rules: 1) keep flying, 2) avoid collisions, and 3) do what the immediate neighbors are doing (Reynolds, 1987; Bialek et al., 2012; Hemelrijk and Hildenbrandt, 2012; Hemelrijk and Hildenbrandt, 2015).

Game theoretic agents can be interacting macromolecules, or pre- and postsynaptic neurons, or neuronal assemblies defined by cofiring neuronal population dynamics. We hypothesize that the utilities that govern the behavior of each neuronal game theoretic agent are bioenergetic functions (Laughlin et al., 1998; Niven et al., 2007), consistent with the idea that by preventing accumulation of reactive oxygen species, action potentials can have a spontaneous, input-independent, metabolic origin when firing rates fall below a homeostatic baseline (Chintaluri and Vogels, 2022). In each case, the system of local, scale-appropriate interactions respects the bioenergetic equilibria that govern their activity. Put another way, dynamical structures (analogous to murmuration configurations) will tend to persist when they correspond to viable signaling conventions. Persistent conventions tend to maintain because like most precedents, they constrain the potential subsequent interactions between agents. This pattern of behavior emerges because most starlings, or biomolecules or neuronal action potential discharge patterns will fail to persistently interact in a manner that is incompatible with the currently persistent conventions — it will simply be too bioenergetically costly to persistently adopt contrarian activities. One can observe this directly in the distribution of cofiring relationships within a population of neurons. The distribution is skewed such that there are many more strongly positive cofiring relationships than anti-cofiring relationships (see Fig. 5). That is however, not to say the anti-cofiring relationships are unimportant. On the contrary, they tend to be rare, their prevalence increases with learning, and they tend to be the most important contributors to the ability to discriminate information from population network activity (Levy et al., 2023).

Once established, persistent conventions tend to further increase their persistence. Consider those agents that are currently behaving in ways that are independent, or even contrary to a currently instantiated convention, for example the neurons with weak synaptic connections that, as a result, discharge independently of the dominant neuronal network pattern. Such neurons may be subsequently recruited to the dominant network discharge pattern. Put another way, they may adopt a discharge pattern that mimics the dominant discharge pattern and then through the costly signaling of synaptic plasticity, become legitimately incorporated into the persistently active network pattern of discharge. One might expect in this case that the weakly cofiring neuron pairs will increase their likelihood to cofire. Although such recruiting of independent neurons into the persistent activity of an established cell assembly appears intuitive, perhaps even routine, an interesting, surprising consequence of this recruitment can occur when the signals of the recruited neurons also provide information that is distinctive from the information that the network processes with its signaling convention. This is in fact observed in the cofiring relationships of hippocampus principal cells under acute intoxication with the psychotomimetic phencyclidine (Kao et al., 2017; Park et al., 2023), or other cognition impairing manipulations, such as tetrodotoxin-induced disinhibition of the contralateral hippocampus (Olypher et al., 2006). In both cases, cell pairs with negative or weak cofiring statistics before the drug manipulation, increase their cofiring under the manipulation, whereas the initially strongly cofiring cell pairs are unchanged. This pattern of selectively increased cofiring coincides with the inability to perform behavioral tasks that require discriminating between relevant and irrelevant information, but the same cofiring increase is not impairing in conditions where the irrelevant environmental information is attenuated (Wesierska et al., 2005; Olypher et al., 2006; Kao et al., 2017). From the perspective of the information that the neuronal activity represents, by adopting the convention, the recruited neurons could have deceived the neuronal network partners that they have been recruited to join. This framework has interesting, and powerful explanatory implications, we will discuss shortly. For the time being, it provides an explanation of observed increase in cofiring of initially anti- and weakly cofiring cell pairs after knowledge impairing manipulations. This game theoretic perspective can also explain another otherwise puzzling observation from ensemble recordings of spatially-tuned neurons in the entorhinal cortex and hippocampal regions that are thought to constitute the brain's navigation system (Fig. 5). We will consider that next.

ENTORHINAL-HIPPOCAMPAL NEURONAL POPULATION DYNAMICS: PHENOMENA IN NEED OF A CONCEPT

We will now elaborate on the entorhinal-hippocampal neuronal system to set the foundation for what we aspire to understand and explain using the concepts of information asymmetric signaling games (Fig. 5(**A**)). We focus on hippocampus neuronal population discharge correlates of spatial information that can be measured by studying the spatial behavior of freely-behaving rodents, as this is our long-standing experimental research program (see Methodology in Appendix 1). In freely-behaving rodent subjects like rats and mice, an environment-specific 20–25% subset of hippocampus principal cells discharge action potentials robustly only when the subject is in cell-specific locations called the cell's firing field (O'Keefe, 1976); place cells are also identified in birds and bats (Ulanovsky and Moss, 2007; Yartsev Michael and Ulanovsky, 2013; Payne et al., 2021). When neurons discharge in this way they are traditionally called "place cells" (Fig. 5(**A,B**)). The medial entorhinal cortex (MEC) contains neurons that signal distance travelled ("grid cells"), head-direction ("head-direction cells") (Sargolini et al., 2006), the presence of environmental borders ("border cells") (Savelli et al., 2008; Solstad et al., 2008) and the current speed of locomotion ("speed cells") (Kropff et al., 2015), with cells tuned to spatial components in other related areas like subiculum and retrosplenial cortex (Lever et al., 2009; Brotons-Mas et al., 2017; Alexander et al., 2020). The 2014 Nobel Prize was awarded for the discoveries of these functional cell classes (Fenton, 2015; Moser et al., 2017). These spatially-tuned MEC neurons project to the hippocampus in multiple, parallel and distinctive pathways, providing multiple sources of the information components for computing "place" (Fig. 5(**A**); Baks-Te Bulte et al., 2005; Canto et al., 2008; Witter, 2006, 2007).

We developed an experimental paradigm in which rats and mice readily navigate on a slowly, continuously rotating circular arena (Fig. 5(**B**) top). The arena rotation dissociates the accessible space into two simultaneous and distinct spatial frameworks. One is stationary defined by room-anchored landmarks, and the other is rotating, defined by arena-anchored stimuli such as scent marks on the rotating surfaces. Animals quickly demonstrate that they understand the room and arena spaces to be distinct, even if the arena never rotates (Fenton et al., 1998; Fenton and Bures, 2003). They demonstrate this knowledge in the active place avoidance paradigm, in response to training during which we punish the animal with a mild electric shock for entering a room-defined zone and/or an arena-defined zone (Fig. 5(**B**) top; Fenton and Bures, 2003;Fenton et al., 1998). In response, they will quickly and selectively avoid the room and/or arena shock zone, which is why we call the behavior two-frame place avoidance. Two-frame place avoidance is one of the most sensitive tasks to disturbed hippocampal function; even inactivating one of the two hippocampi makes the animals unable to learn, consolidate, or remember the location of the shock zone (Cimadevilla et al., 2000; Cimadevilla et al., 2001; Kubik and Fenton, 2005; Wesierska et al., 2005; Kelemen and Fenton, 2010).

We investigated whether during two-frame place avoidance, the spatially-tuned cells in hippocampus and entorhinal cortex would exhibit spatial tuning in the room frame or the arena frame. Despite the animal navigating the rotating environment extremely well, very few neurons (a fraction of a percent) demonstrate their place cell, grid cell, or head-direction cell spatial-tuning properties during rotation; the tuning returns once the rotation stops (Fig. 5(**B**)). We have shown that this apparent loss of spatial tuning is because the neuronal populations collectively signal the animal's location and direction in an internally-organized, multistable manner such that the ensemble activity patterns switch between representing the current location and direction in either the room frame or the arena frame, but not both (Kelemen and Fenton, 2010; Kelemen and Fenton, 2013; Talbot et al., 2018; van Dijk and Fenton, 2018; Park et al., 2019; Chung et al., 2021). Multistable switching between room and arena representations is rapid (sub-second) and occurs in a purposeful way with a period in the range of 10 s (Kelemen and Fenton, 2010; Kelemen and Fenton, 2013; van Dijk and Fenton, 2018; Park et al., 2019).

These multistable population discharge dynamics in maintained environmental conditions indicate that neuronal activity is internally-organized and variably registered to the environment in a manner that takes into account the distinctive spatial frames and registers neuronal ensemble activity to the spatial frame that is more currently useful (i.e., use the room frame when near the stationary room-frame shock zone and use the arena frame when near the rotating, arena shock zone) (Kelemen and Fenton, 2010; Kelemen and Fenton, 2013; van Dijk and Fenton, 2018; Park et al., 2019). The cofiring relationships of simultaneously recorded cell pairs provides additional, independent evidence for this strong internal organization of neuronal discharge patterns. This conclusion can be arrived at by computing the pairwise correlations among all simultaneously recorded cell pairs (Schneidman et al., 2006). We observe that the pairwise coactivity measures of their correspondence is indistinguishable during recordings with the arena stable compared to recordings with the arena rotating. An example from the head-direction cells recorded from superficial layers of the medial entorhinal cortex is shown in Fig. 5(**C**) top, indicating that the population discharge of these ‘‘same-function" neurons is dynamically rigid (stationary and steady) (Park et al., 2019). A similarly persistent set of cofiring relationships is also observed for all the simultaneously recorded neurons in superficial MEC (Fig. 5(**C**) middle), further indicating that despite mixed tuning to different spatial variables (Park et al., 2019), as a whole, the MEC network manifests the stationary and steady temporal discharge dynamics that are characteristic of neuronal attractor dynamics (Yoon et al., 2013; Chaudhuri et al., 2019). Like MEC, the population discharge properties of hippocampus also exhibit attractor dynamic properties, in that the collective activity of MEC cells, or the collective activity of the cells of the hippocampus CA3 or CA1 subfields tend to exist as relatively stable patterns of activity that are readily described as a low-dimensional manifold in the high-dimensional activity space that is defined by the independent activities of the population of cells (Samsonovich and McNaughton, 1997; McNaughton et al., 2006; Yoon et al., 2013; Chaudhuri et al., 2019; Gardner et al., 2021; Nieh et al., 2021). Evidence of these low-dimensional manifolds of neuronal activity can be readily measured as conserved pairwise coactivity patterns across the network, which are known to represent higher-order correlations (Fig. 5(**C**); Schneidman et al., 2006; Levy et al., 2023).

DO SIGNALING GAMES OFFER ANYTHING THAT INFORMATION MAXIMIZATION DOESN'T ALREADY?

Our observations of multistable frame-specific positional and directional population discharge, and persistence of internally-organized cofiring discharge relationships (Kelemen and Fenton, 2010; Talbot et al., 2018; van Dijk and Fenton, 2018; Park et al., 2019) can be readily accommodated by the information maximization (Infomax) perspective that is a dominant conceptualization of how information processing is serially organized across connected neuronal networks with related input–output functions (Bell and Sejnowski, 1997; Linsker, 1997), such as the MEC and hippocampal computations of environmental space (Solstad et al., 2006).

Neuronal ensemble discharge patterns are assumed to be the neuronal representations and/or instantiations of mental objects, in short, the expression of cognitive phenotypes. This view assumes that generating reliable representations of cognitive variables essentially constitutes proper entorhinal-hippocampal ensemble function. Such assertions emerge from the Infomax hypothesis that has been important for systems neuroscience because it offers a unifying principle (Barlow, 1961). Accordingly, neuronal computations at each level of the nervous system would operate to maximize the mutual information between inputs and outputs that the computation is operating on, and so couple the information content across a series of such neuronal computations, the output of one serving the input to the next. While Infomax principles are readily applied to predict what neuronal discharge might signal, it is difficult to apply these principles to the biochemistry and cell biology that operates and underlies the neuronal discharge that define the mental objects of interest, in particular when a cell's discharge can also have a metabolic origin (Chintaluri and Vogels, 2022). Even in straightforward terms of information representation as it concerns the MEC – hippocampal spatial information processing system, Infomax concepts struggle to account for the lack of, and/or transient effects of inhibitory designer receptor exclusively activated by designer drugs (DREADD) and optogenetic silencing of MEC, and even the effects of MEC lesion on downstream hippocampal place cells, despite effects of excitatory stimulation (Miao et al., 2015; Rueckemann et al., 2016; Kanter et al., 2017; Schlesiger et al., 2018).

Furthermore, as it concerns the place avoidance experiments on a rotating arena, Infomax concepts are severely challenged to explain our additional observation that the cofiring relationships between simultaneously-recorded hippocampal and MEC neurons (Fig. 5(**C**) bottom) is only weakly persistent across the stable and rotating arena conditions, despite both the entorhinal (Park et al., 2019) and the hippocampal populations each exhibiting strongly persistent, attractor-like cofiring relationships (Levy et al., 2023). If MEC representations of space project to hippocampus then their discharge should be informationally coupled, and the discharge representations should be informationally coherent, perhaps even increasing in fidelity according to Infomax predictions. But as Fig. 5(**C**) suggests, the discharge coupling is unstable, and the frame-specific multistability in hippocampus (Kelemen and Fenton, 2010; van Dijk and Fenton, 2018) is observed to have a different, weakly opposite relationship to frame-specific behavior as is observed for MEC frame-specific multistability (Park et al., 2019). Because it is difficult to reconcile these observations within the Infomax framework, we were motivated to seek an alternative. We contend that the framework of signaling games has utility for understanding these challenging features of hippocampal neuronal population dynamics at the level of systems neuroscience, as well as at other levels of biological organization, including cell biology and biochemical signaling that others have demonstrated (Smith, 2000; Jee et al., 2013).

The game theoretic framework of fundamentally coupled senders and receivers provides a natural analytical language and framework for characterizing these neuronal dynamics and for analyzing their multiple steady states as game theoretic Nash equilibria. Indeed, the manifold patterns of attractor-like neuronal activity (Amit, 1992) are readily described by the concept of ‘‘cellularization” that naturally emerges in the game theoretic language. Through reliable and oftentimes costly signaling like metabolically-instantiated synaptic plasticity, multiple game theoretic agents, in this case neurons within an interconnected network, can settle into stable population equilibrium states that are characterized by signaling conventions within the network acting as a single collective. In other words, robust neuronal population dynamics produce activity patterns that themselves act as game theoretic senders and/or receivers, adopting the form of a dynamically ‘‘cellularized” object, despite being composed of many individual neuron components. Note that this process of cellularization by adoption of a set of signaling conventions naturally defines cross-scale interactions and information transfer that are seamlessly transcended by the game theoretic formalism, whereas such cross-scale analysis have been a challenge for other frameworks and conceptualizations. In the next section we consider the entorhinal cortex and hippocampus, each as a cellularized signaling object comprised of millions of neurons, the hippocampus collectively signaling positions for example, and the entorhinal cortex collectively signaling components of space, for example. Entorhinal – hippocampal interactions can then be characterized and analysed as game theoretic sender-receiver interactions, despite the reality that these interactions are implemented by individual neurons in the two areas interacting as single neurons do.

ENTORHINAL-HIPPOCAMPAL NEUROBIOLOGY AS AN INFORMATION ASYMMETRIC SIGNALING GAME

We believe that the ideas of information asymmetric signaling games can readily and advantageously apply to neuronal systems. To that end we now reconceptualize the neurobiological interactions that were described above in these game theoretic terms. Simple conserved game conventions can first arise from arbitrary dependencies, much like how spuriously causally coactive neurons can undergo synaptic plasticity so they are more likely to cofire in the future, and in so doing establish a neuronal firing sequence convention. Selection for useful conventions can then privilege those with utility so they persist with potentially increasing complexity, much like standard reinforcement learning models and Hebbian learning rules assert.

It may be difficult to appreciate how multi-stable cognitive organization can emerge from agents interacting locally in simple games. It is therefore important to emphasize how slight deviations in initial conditions can create lasting differences in evolving dynamical systems. For instance, in area CA1 of the hippocampus, multiple representations of the same spatial environment can coexist as recurring stable patterns of firing (Sheintuch et al., 2020). However, with time and additional exposure to the environment, these representations tend to deviate, according to the signaling game framework, because the information asymmetry among the representations generates persistent distinctions.

Mechanisms to persist, to maintain persistence and to increase complexity of signaling conventions typically require metabolic and other resources and so come at a cost that most agents will not engage (Niven et al., 2007). Such exclusions of non-compatible agents increases the reliability of the signaling system as a whole, which can make costly signaling advantageous (de Ruyter van Steveninck and Laughlin, 1996). Synaptic plasticity is metabolically costly, and molecular biosynthesis needs to be coordinated pre- and postsynaptically for the structural plasticity that changes the shapes and sizes of synapses engaged during learning. This coordination results in a myriad of local processes including actin polymerization, synthesis and then translocation of synaptic proteins (Chen et al., 2022). In addition, once strengthened, potentiated synapses will result in greater trans-membrane ion currents and thus dissipation of the sodium and potassium ionic concentration gradients, which, to be restored, will require increased Na/K ATPase biosynthesis and activity. Indeed, the upregulation of PKMζ, which is necessary and sufficient for the long-term maintenance of wildtype long-term potentiation at hippocampal and other synapses, coincides with the upregulation of the Na/K ATPase (Tian et al., 2008). This observation is consistent with the notion that the increased communication comes at an energetic cost and so once in place, is likely to contribute to the information asymmetry.

As mentioned previously, information symmetric games are those in which both players reveal complete information about each other and consequently are confident about how the other rational players will strategically respond in each scenario. Information asymmetry occurs when one player (the sender) possesses information about its signaling type that is not available to the other player (the receiver). By introducing information asymmetry, we can better describe molecular signaling cascades and neuronal networks. As an example of a cascade, consider the multiple roles of intracellular calcium, which can initiate postsynaptic vesicle release, or changes in gene expression, or programmed cell death, each depending on the origin of the calcium. In the case of a neuronal network, considering that each neuron receives $\sim 10^4$ distinct synaptic inputs, the consequences of their activation are integrated to change the transmembrane potential at the remote location of the axon hillock. The neuron's axon hillock is uncertain about which subset of the $\sim 10^4$ synapses activated in temporal and spatial summation, causing it to generate an action potential. This fundamental uncertainty creates information asymmetry and the opportunity for game theoretic deception.

Game theoretic deception occurs when there is information asymmetry and insufficient common interest between the sender and receiver. Insufficiently costly signaling conventions may be exploited by deceptive agents exploiting the convention, even though they degrade the reliability of signaling. Examples of these preserved conventions include canonical neurotransmitters and receptors, cell-specific signaling proteins, and immediate early genes. Examples of deceptive agents include but are not limited to pharmacological or natural agents that mimic the biomolecular structure of these conventional signals. These conventions are ubiquitous across vertebrates and can be found functioning in both *in vitro* slice preparations as well as in *in vivo* preparations such as recordings from awake freely-behaving subjects. Essentially the same local cell-biological biochemical constraints underlie the structural dynamics of all neuronal populations during transcription, translation, membrane depolarization, action potential generation, and synaptic and neuronal population synchrony that is measured as oscillations in local field potentials and dynamic cofiring patterns in neuronal ensemble activity. While not exhaustive, we believe these examples illustrate there is substantial and natural potential to translate neurobiological phenomena and theory into the formal and descriptive language of information asymmetric signaling game theory. In fact, as we alluded to above, a game theoretic framework offers mechanistic accounts for neuronal network phenomena that are otherwise difficult to explain, as we will now illustrate, by applying the concept of deception.

DECEPTION: INFORMATION INCOMPLETENESS VS INFOMAX

Within a signaling game, the receiver's signaling strategies is defined by the ensemble of responses of a particular receiver with respect to a given set of signals. When senders can flexibly alter their signaling in response to the subjective probabilities of other agents, their signals can cause the receiver to respond in ways that benefit the sender at the expense of the receiver. This phenomenon is understood within the signaling game framework as deception, whereby subjective probabilities of receivers are leveraged by senders at the expense of the receivers utilities. When deception occurs, conventions are destabilized because signals are de-correlated from the receiver's expected utility. When the interests of senders and receivers are adversarially structured, deception results in increased utility for the sender at the expense of the receiver. Under some conditions, this could generate a destabilizing feedback loop that may lead to the breakdown of signaling, which could be clinically relevant in degenerative neuropathology. Under different conditions more favorable for cognition, destabilizing signals could result in more robust conventions flexible to distortion, deception, and new information, which may be relevant for learning and memory. We stress that in a signaling game ''deception'' doesn't require an epistemic agent, as in the well-known cases of frogs and snakes (Casey et al., 2021) but also that ''utility'' in a signaling game is not the same as ''utility'' in natural selection, though one can lead to the other. For example, winner-take-all neuronal network configurations in which neurons compete for inputs and coactivity in output representations is an example of competitive signaling game utility that results in output contrast (Fig. 4). Distinctive output can be a net positive adaptive advantage for the organism as a whole as we have observed in hippocampal responses to cognitive control training (Dvorak et al., 2018; Chung et al., 2021; Dvorak et al., 2021), or in the case of pathology a net negative as we have observed with the psychotomimetic agent phencyclidine (Kao et al., 2017; Park et al., 2023).

Let's consider the notion of game theoretic deception in the context of an explicit neuronal discharge phenomenon that we have studied in the neuronal dynamics of the hippocampus. Hippocampal neurons, residing in the CA1 output subfield, can discharge as if to represent the current location of a mouse. This interpretation of the observables leads one to think of the ''place cell'' phenomenon. From there it is relatively straightforward to understand how a place cell network of neuronal activity is ''encoding'' the current location. However, we also observe that, at least transiently, the

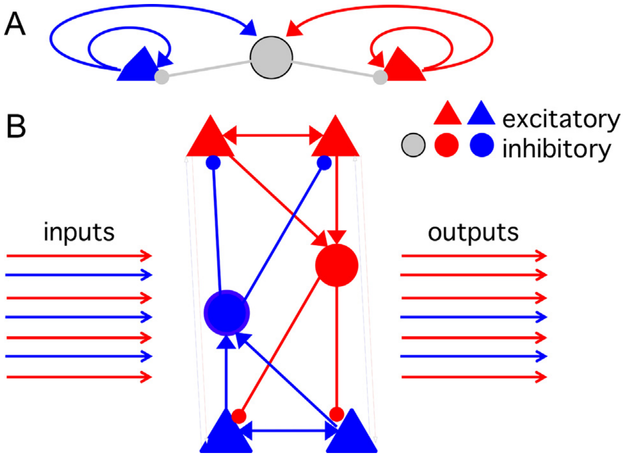

Fig. 4. Winner-take-all competitive networks function to select the strongest response amongst a number of competing responses. (**A**) Many network architectures have winner-take-all properties, but the most basic motif is to have sufficient mutual excitation amongst the same-function network elements and mutually-inhibitory coupling between the excitatory elements. (**B**) Illustration of a competitive network where each node represents a population of the nodes. Such a system will transform a set of inputs of varied type into an output that is dominated by type of the strongest inputs.

same CA1 neuronal network's discharge will not signal the current location (Dvorak et al., 2018). Instead for about 500 ms, the discharge will represent a distant location that, during a spatial memory task, can be shown to represent the mouse's recollection of a remote goal or some other behaviorally-important location (Fig. 6). Conventional explanations do not easily explain how such non-local place cell activity can come about, and the field has hypothesized that there might be specialized "goal" or "reward" cells, rather than place cells that sometimes act like goal or other types of functionally-defined cells (Poucet and Hok, 2017; Gauthier and Tank, 2018; Duvelle et al., 2019). Recent work shows that any place cells can transiently express non-local place cell discharge that otherwise resembles routine locally-legitimate place cell firing. Because the CA1 network behavior has switched from signaling local positions to signaling remote locations, what in ordinary neuroscientific parlance is context-dependent activity, in game theoretic terms, the network has been deceived into signaling a remote location, similar to Player II in the UNCERTAINTY game we considered in Fig. 1. The deception appears especially strong and maladaptive in *Fmr1-null* mutant mice that model the genetic defect in human Fragile X Syndrome, the leading cause of intellectual disability and autism. *Fmr1-null* mutant mice express synaptic protein expression dysregulation (Broek et al., 2015; Thomson et al., 2017) and as a consequence of memory training, excessive synaptic plasticity in hippocampus CA1 (Talbot et al., 2018). When challenged with a novel location after learning an initial location, *Fmr1-null* CA1 place cell network discharge and the mouse's behavior both recollect the formerly correct location rather than the currently correct location for the memory task (Dvorak et al., 2018). These phenomena are straightforward to explain as instances of deception within asymmetric signaling games, and the neuronal mechanism for transiently switching the hippocampus information processing from encoding to recollection is consistent with this game theoretic explanation (Dvorak et al., 2021). There is a transient, strong synchronous discharge event that originates in the upstream MEC and is observed in the dentate gyrus (DG). The event is called a MEC-originating dentate spike (DS_M). The DS_M acts like a control switch for the MEC → DG → CA3 → CA1 neuronal circuit (Fig. 6) that changes information processing in the circuit in a way that causes the CA3 cells to discharge in a manner that is informationally disconnected from the ongoing location-specific discharge in the dentate gyrus. From the game theoretic perspective, the DS_M-triggered CA3 activity has deceived CA1 into discharging in a manner that appears normal, only, instead of representing the current location, the DS_M-associated CA1 discharge represents a remote, recollected location (Fig. 6). A similar, non-local place cell ensemble discharge phenomenon has also been described that is easily explained in game theoretic terms by this deception concept. Again, on a subsecond timescale, the ensemble discharge of CA1 neurons will toggle between representing non-local places that a rat may visit in the near future (Kay et al., 2020). It is as if the activity during that moment represents something akin to the rat's aspirations or intentions rather than where it is, which is not easy to explain from a feedforward information maximization conceptualization of hippocampal spatial information processing but is a natural consequence of deceptive signaling.

WRAPPING UP

We have recognized that signaling game theory can successfully explain, describe, and model complex phenomena in diverse fields ranging from biomolecular evolution to economics. From the small-scale to the large-scale, at each level of analysis there are biological signaling agents transmitting signals in attempts to maximize their respective local utility. Information asymmetry between signaling agents can lead to stable equilibria under cooperative circumstances, equilibria which may persist due to the low cost of signaling relative to the acquired utility of signaling for the local agents. These equilibria might be cognitively beneficial or deleterious depending on the context, and factors influencing these equilibria are under the regulatory influence of natural selection. Persistent signaling equilibria are considered conventions, and many conventions are preserved across species, adding credence to the universality of signaling systems and their emergent conventions as neurobiological objects of study.

Neurons use biochemical and bioelectric signals to differentially allocate metabolic resources to particular "privileged" connections. Although the local metabolic constraints of neurons are determined by biochemical considerations, natural selection and experience can act on these asymmetries at the level of their emergent features. These emergent features are phenotypes, and for our consideration these are the phenotypes of cognition and behavior, which are correlates of neuronal ensemble discharge patterns. Coordinating behavior into coherent patterns is a

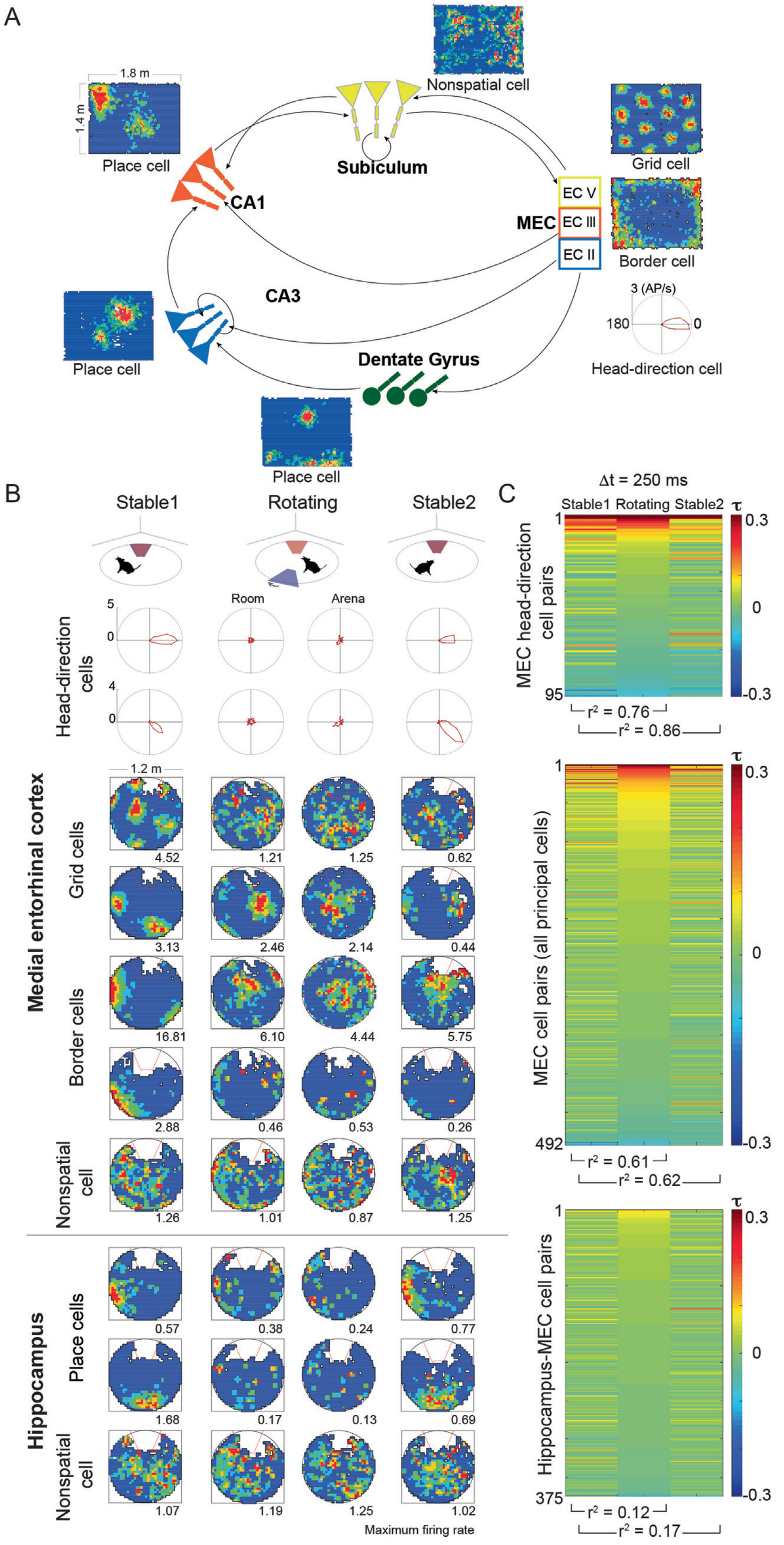

Fig. 5. Hippocampal network functional architecture for spatial cognition, and functional cellularization. (**A**) schematic illustrating the functional connectivity amongst the distinct subregions of the medial entorhinal cortex and the hippocampus. Only excitatory inputs are illustrated, emphasizing the interareal connections, whereas the inhibitory connections tend to be mostly local (intra-areal) and have been omitted. Note that the innervations target distinct dendritic compartments. The spatial firing properties of principal cells in each area are also indicated by an example session-averaged firing rate map from neuronal recordings while a rat explored a rectangular arena. Note how grid, head direction, and border spatially-tuned cells of MEC represent component variables from which place can be computed (distance, azimuthal direction, and environmental boundary, respectively) from the spatially-tuned MEC inputs to the subfields in which place cell spatial firing patterns are the most frequently observed spatial firing patterns. (**B**) Top-to-bottom: Schematic of stable-rotating-stable experimental active place avoidance experimental conditions. Example polar firing rate representations of two head-direction cells and blue-to-red color-coded firing rate maps illustrate the typical spatial tuning of MEC and hippocampus (CA1) cells recorded while rats navigate during two-frame place avoidance in a stable-rotating-stable triad of 30-min recordings. The rotating session dissociates the accessible space into two spatial frames, a stationary room frame and a rotating arena frame, and the two frame-specific firing rate maps are provided for each cell. The number under each map is the minimum rate in the red category measured in AP/s units. Note how the spatial tuning observed during the stable sessions is lost during the rotation. (**C**) Distribution of cell pair cofiring measured as Kendall's correlation (τ) computed at 250 ms resolution. Each simultaneously-recorded pair of cells was recorded in the stable-rotating-stable triad session. The cell pairs are sorted in descending order by the value of τ during the rotating recording and the order of cell pairs is maintained for all three recordings. Note (1) that the correlation patterns skew to significant positive values with few significant negative values, and (2) that the correlation patterns strongly persist across the session triad (r^2 coefficient of determination values given below the plots) for MEC recordings. In contrast, the persistence of cofiring is much weaker for MEC-hippocampus cell pairs (Stable – Stable: MEC-MEC pairs ($r = 0.79$) vs. MEC-HPC pairs ($r = 0.41$) $z = 9.24$, $p \sim 0$; Stable-Rotating: MEC-MEC pairs ($r = 0.78$) vs. MEC-HPC pairs ($r = 0.35$) $z = 9.88$, $p \sim 0$). See Appendix 1 for Methods.

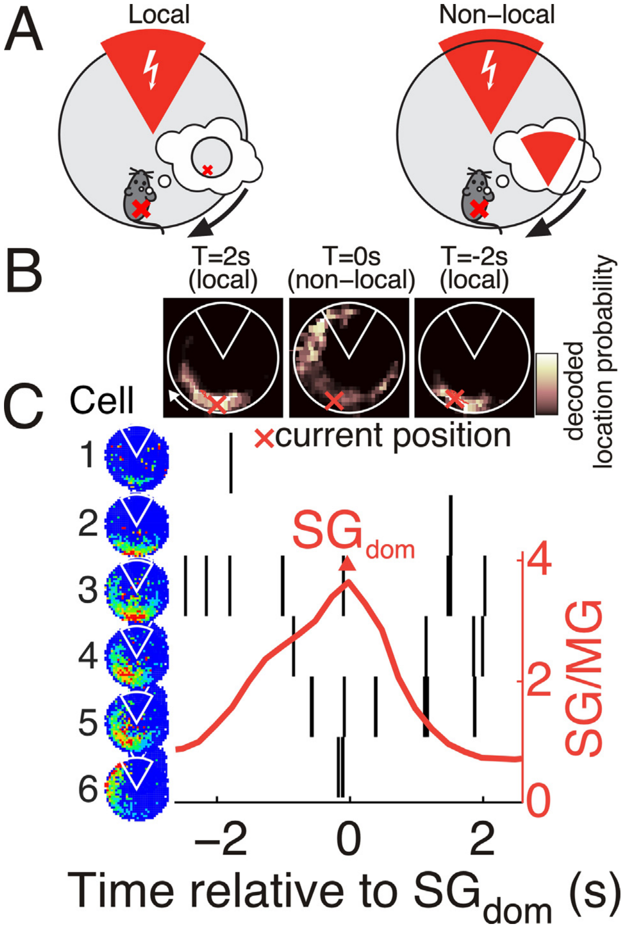

Fig. 6. Non-local hippocampal network discharge as deception. (**A**) Cartoon illustrating that hippocampus place cell activity can typically be decoded to accurately identify the subject's current position (left). However, at times the position decoding is non-local in that it points to a position remote from the subject's current position (right). This non-local decoding is often to the vicinity of a shock zone (red sector) if the mouse has been trained to avoid that shock zone. Furthermore, the non-local representation of position tends to occur ~ 2 seconds before the subject will make avoidance movements away from the shock zone, suggesting that it is a recollection of the shock zone's location. (**B**) Bayesian posterior maps of the likelihood of a mouse's position calculated on the basis of a 500-ms observation of place cell population discharge centered at the times indicated. (**C**) T = 0 s is the moment that slow gamma dominance was observed in the concurrent CA1 local field potential (SGdom). SGdom is computed as the ratio of slow (~30 Hz) to mid-frequency (~70 Hz) gamma (MG) in the local field potential, and T = 0 is marked as the peak of the sufficiently positive SG/MG ratio. The firing rate maps of six place cells are shown ordered from the one with the firing field closest to the mouse's current location to the one with the firing field distant from the current location and near the shock zone. Note that during this 5-s data segment, that despite the mouse moving very little, which place cells discharge action potentials (rasters shown) changes from the cells with fields at the current location, to the cells with fields near the shock zone, remote from the mouse. It is not obvious what sensory cues would switch firing from representing local position to non-local position, but SGdom transiently switching CA1 network function from processing local place information based on sensory cues to processing non-local information based on memory predicts these observations. SGdom is triggered by a medial entorhinal cortex-originating dentate spike event that switches hippocampal information processing from sensation-based encoding to memory based recollection (Dvorak et al., 2021). Figure based on (Dvorak et al., 2018).

unifying feature of all evolutionary signaling systems, and the reliability of signaling is determined by how well a signal can predict a sender's type. For instance, a signaling system approximating Infomax consists of senders reliably signaling their type, thereby reducing distortion, and generating optimal and predictable responses. But, as mentioned above, observable, and adaptive hippocampal function demands more than optimally reliable information. Spatial information maximization is instead merely one end in a spectrum of signaling organization: the optimally informative end. On the other end of the spectrum is completely uninformative signaling (so-called babbling equilibria) whereby signals do not reliably distinguish among sender types. Either end of this spectrum alone is untenable for the demands on hippocampal function. It is therefore our contention that entorhinal-hippocampal organization can be understood as the cellularization of scale-specific signaling agents ranging from genes, proteins, neurons, and neuronal populations that interact according to shared signaling conventions. Every cellularized group of signaling agents is organized somewhere along the babbling versus separating signaling spectrum. From the scale of neuronal ensemble dynamics, and emergent sensorimotor features, information-rich separating equilibria can approximate information maximization paradigms and attractor networks which have been highly fruitful to systems neuroscience research. These populations of highly cofiring neurons are well suited to encode reliable sensorimotor features of the world, allowing an organism to benefit from the high degree of cellularization and reliable signaling. Furthermore, signaling game theory also provides an explanation for aspects of hippocampal activity which are not well understood from a systems level information maximization paradigm. Populations of cofiring neurons further from separating equilibria generate less cellularization, allowing signaling conventions to adapt to changing relations between the organism and the world. We observe that loose correlations observed in less cellularized populations underlie *Fmr1*-null mutant CA1 hippocampus' susceptibility to deception (Talbot et al., 2018). Moreover, the dynamic restructuring of population vector correlations observed during remapping experiments in area CA1 (Redish et al., 2000; Jackson and Redish, 2007; Kubie et al., 2020; Levy et al., 2023) provide further support for the significance of signaling equilibria that can transiently stray from separating equilibria to establish new conventions. In summary, the tension generated between different cellularized agents and their respective level of cellularization can account for the complex array of functions attributed to the hippocampus, functions which emerge from metabolic and energetic constraints of individual neurons and scale up to population-level phenotypes of cognition and memory.

ACKNOWLEDGEMENTS

We are grateful to Will Casey, Cristina Savin and Kshaunish Soni, and Mike Wigler for discussions.

SUPPORT

This work was supported by the National Institutes of Health grants R01NS105472, R01MH115304.

CONTRIBUTIONS

Research conceptualization AAF, BM, JB, JH; Data collection and analysis EHP; Figure creation AAF, EHP; AAF wrote the paper with contributions from BM, JB, JH.

DECLARATIONS OF INTEREST

None.

REFERENCES

Alexander AS, Robinson JC, Dannenberg H, Kinsky NR, Levy SJ, Mau W, Chapman GW, Sullivan DW, et al. (2020) Neurophysiological coding of space and time in the hippocampus, entorhinal cortex, and retrosplenial cortex. Brain Neurosci Adv 4 2398212820972871.

Amit DJ (1992) Modeling brain function. Cambridge: Cambridge University Press.

Andersen BP (2022) Autistic-like traits and positive schizotypy as diametric specializations of the predictive mind. Perspect Psychol Sci 17:1653–1672.

Baks-Te Bulte L, Wouterlood FG, Vinkenoog M, Witter MP (2005) Entorhinal projections terminate onto principal neurons and interneurons in the subiculum: a quantitative electron microscopical analysis in the rat. Neuroscience 136:729–739.

Barlow HB (1961) Possible principles underlying the transformations of sensory messages. In: Rosenblith W, editor. Sensory Communication. Cambridge, MA: M.I.T. Press. p. 217–234.

Bassett DS, Sporns O (2017) Network neuroscience. Nat Neurosci 20:353–364.

Bell AJ (2008) Towards a cross-level theory of neural learning. 27th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, pp. 56–73.

Bell AJ, Sejnowski TJ (1997) The ‘‘independent components” of natural scenes are edge filters. Vision Res 37:3327–3338.

Bialek W, Cavagna A, Giardina I, Mora T, Silvestri E, Viale M, Walczak AM (2012) Statistical mechanics for natural flocks of birds. Proc Natl Acad Sci 109:4786.

Binmore K (2007) Game Theory: A Very Short Introduction. OUP Catalogue. Oxford University Press.

Bouchacourt D, Baroni M (2018) How agents see things: On visual representations in an emergent language game. Brussels, Belgium: Association for Computational Linguistics. p. 981–985.

Broek JAC, Lin Z, Van't Spijker H, Ozcan S, De Gruiter HM, Haasdijk ED, Willemsen R, De Zeeuw CI, et al. (2015) Proteomics investigation identifies prominent changes in synapse-related proteins in a fragile X mouse model. BMC Neurosci 16:P21.

Brotons-Mas JR, Schaffelhofer S, Guger C, O'Mara SM, Sanchez-Vives MV (2017) Heterogeneous spatial representation by different subpopulations of neurons in the subiculum. Neuroscience 343:174–189.

Canto CB, Wouterlood FG, Witter MP (2008) What does the anatomical organization of the entorhinal cortex tell us? Neural Plast 2008 381243.

Casey W, Kellner A, Memarmoshrefi P, Morales JA, Mishra B (2019) Deception, identity, and security: the game theory of Sybil attacks. Commun ACM 62:85–93.

Casey W, Massey SE, Mishra B (2020) How signaling games explain mimicry at many levels: from viral epidemiology to human sociology. Res Sq.

Casey W, Massey SE, Mishra B (2021) How signalling games explain mimicry at many levels: from viral epidemiology to human sociology. J R Soc Interface 18:20200689.

Chaudhuri R, Gercek B, Pandey B, Peyrache A, Fiete I (2019) The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. Nat Neurosci 22:1512–1520.

Chen N, Zhang Y, Adel M, Kuklin EA, Reed ML, Mardovin JD, Bakthavachalu B, VijayRaghavan K, et al. (2022) Local translation provides the asymmetric distribution of CaMKII required for associative memory formation. Curr Biol 32:2730–2738.e2735.

Chintaluri C, Vogels TP (2022) Metabolically driven action potentials serve neuronal energy homeostasis and protect from reactive oxygen species. bioRxiv. 2022.2010.2016.512428.

Chung A, Jou C, Grau-Perales A, Levy ERJ, Dvorak D, Hussain N, Fenton AA (2021) Cognitive control persistently enhances hippocampal information processing. Nature 600:484–488.

Cimadevilla JM, Fenton AA, Bures J (2000) Functional inactivation of dorsal hippocampus impairs active place avoidance in rats. Neurosci Lett 285:53–56.

Cimadevilla JM, Wesierska M, Fenton AA, Bures J (2001) Inactivating one hippocampus impairs avoidance of a stable room-defined place during dissociation of arena cues from room cues by rotation of the arena. PNAS 98:3531–3536.

Crawford VP, Sobel J (1982) Strategic Information Transmission. Econometrica 50:1431–1451.

de Almeida L, Idiart M, Lisman JE (2009) A second function of gamma frequency oscillations: an E%-max winner-take-all mechanism selects which cells fire. J Neurosci 29:7497–7503.

de Ruyter van Steveninck RR, Laughlin SB (1996) The rate of information transfer at graded-potential synapses. Nature 379:642–645.

DeDeo S, Krakauer DC, Flack JC (2010) Inductive game theory and the dynamics of animal conflict. PLoS Comput Biol 6 e1000782.

Dongen SV (2006) Fluctuating asymmetry and developmental instability in evolutionary biology: past, present and future. J Evol Biol 19:1727–1743.

Duvelle É, Grieves RM, Hok V, Poucet B, Arleo A, Jeffery KJ, Save E (2019) Insensitivity of place cells to the value of spatial goals in a two-choice flexible navigation task. J Neurosci 39:2522.

Dvorak D, Chung A, Park EH, Fenton AA (2021) Dentate spikes and external control of hippocampal function. Cell Rep 36 109497.

Dvorak D, Radwan B, Sparks FT, Talbot ZN, Fenton AA (2018) Control of recollection by slow gamma dominating mid-frequency gamma in hippocampus CA1. PLoS Biol 16.

Fenton AA (2015) Coordinating with the ‘‘Inner GPS”. Hippocampus 25:763–769.

Fenton AA, Bures J (2003) Navigation in the moving world. In: Jeffery K, editor. The Neurobiology of Spatial Behaviour. Oxford: Oxford University Press.

Fenton AA, Wesierska M, Kaminsky Y, Bures J (1998) Both here and there: simultaneous expression of autonomous spatial memories in rats. PNAS 95:11493–11498.

Fruchart M, Hanai R, Littlewood PB, Vitelli V (2021) Non-reciprocal phase transitions. Nature 592:363–369.

Gardner RJ, Hermansen E, Pachitariu M, Burak Y, Baas NA, Dunn BA, Moser M-B, Moser EI (2021) Toroidal topology of population activity in grid cells. bioRxiv. 2021.2002.2025.432776.

Gauthier JL, Tank DW (2018) A dedicated population for reward coding in the hippocampus. Neuron 99:179–193. e177.

Guzowski JF, Knierim JJ, Moser EI (2004) Ensemble dynamics of hippocampal regions CA3 and CA1. Neuron 44:581–584.

Hemelrijk CK, Hildenbrandt H (2012) Schools of fish and flocks of birds: their shape and internal structure by self-organization. Interface Focus 2:726–737.

Hemelrijk CK, Hildenbrandt H (2015) Scale-free correlations, influential neighbours and speed control in flocks of birds. J Stat Phys 158:563–578.

Hopfield JJ, Tank DW (1985) ‘‘Neural” computation of decisions in optimization problems. Biol Cybern 52:141–152.

Hsieh C, Tsokas P, Grau-Perales A, Lesburguères E, Bukai J, Khanna K, Chorny J, Chung A, et al. (2021) Persistent increases of PKMζ in memory-activated neurons trace LTP maintenance during spatial long-term memory storage. Eur J Neurosci.

Jackson J, Redish AD (2007) Network dynamics of hippocampal cell-assemblies resemble multiple spatial maps within single tasks. Hippocampus 17:1209–1229.

Jee J, Sundstrom A, Massey SE, Mishra B (2013) What can information-asymmetric games tell us about the context of Crick's 'frozen accident'? J R Soc Interface 10:20130614.

Jenkins A, Zhu L, Hsu M (2016) Cognitive neuroscience of honesty and deception: A signaling framework. Curr Opin Behav Sci 11:130–137.

Jonas E, Kording KP (2017) Could a Neuroscientist Understand a Microprocessor? PLoS Comput Biol 13.

Kanter BR, Lykken CM, Avesar D, Weible A, Dickinson J, Dunn B, Borgesius NZ, Roudi Y, et al. (2017) A novel mechanism for the grid-to-place cell transformation revealed by transgenic depolarization of medial entorhinal cortex layer II. Neuron 93:1480–1492. e1486.

Kao HY, Dvorak D, Park E, Kenney J, Kelemen E, Fenton AA (2017) Phencyclidine discoordinates hippocampal network activity but not place fields. J Neurosci 37:12031–12049.

Kaski S, Kohonen T (1994) Winner-take-all networks for physiological models of competitive learning. Neural Netw 7:973–984.

Kay K, Chung JE, Sosa M, Schor JS, Karlsson MP, Larkin MC, Liu DF, Frank LM (2020) Constant sub-second cycling between representations of possible futures in the hippocampus. Cell.

Kelemen E, Fenton AA (2010) Dynamic grouping of hippocampal neural activity during cognitive control of two spatial frames. PLoS Biol 8 e1000403.

Kelemen E, Fenton AA (2013) Key features of human episodic recollection in the cross-episode retrieval of rat hippocampus representations of space. PLoS Biol 11 e1001607.

Knierim JJ (2002) Dynamic interactions between local surface cues, distal landmarks, and intrinsic circuitry in hippocampal place cells. J Neurosci 22:6254–6264.

Kropff E, Carmichael JE, Moser MB, Moser EI (2015) Speed cells in the medial entorhinal cortex. Nature 523:419–424.

Kubie JL, Levy ERJ, Fenton AA (2020) Is hippocampal remapping the physiological basis for context? Hippocampus 30:851–864.

Kubik S, Fenton AA (2005) Behavioral evidence that segregation and representation are dissociable hippocampal functions. J Neurosci 25:9205–9212.

Laughlin SB, de Ruyter van Steveninck RR, Anderson JC (1998) The metabolic cost of neural information. Nat Neurosci 1:36–41.

Leutgeb S, Leutgeb JK, Treves A, Moser MB, Moser EI (2004) Distinct ensemble codes in hippocampal areas CA3 and CA1. Science 305:1295–1298.

Lever C, Burton S, Jeewajee A, O'Keefe J, Burgess N (2009) Boundary vector cells in the subiculum of the hippocampal formation. J Neurosci 29:9771–9777.

Lever C, Wills T, Cacucci F, Burgess N, O'Keefe J (2002) Long-term plasticity in hippocampal place-cell representation of environmental geometry. Nature 416:90–94.

Levy ERJ, Carrillo-Segura S, Park EH, Redman WT, Hurtado J, Chung S, Fenton AA (2023) A manifold neural population code for space in hippocampal coactivity dynamics independent of place fields. Cell Reports: 2021.2007.2026.453856.

Linsker R (1997) A local learning rule that enables information maximization for arbitrary input distributions. Neural Comput 9:1661–1665.

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

Massey SE, Mishra B (2018) Origin of biomolecular games: deception and molecular evolution. J R Soc Interface 15:20180429.

McNaughton BL, Battaglia FP, Jensen O, Moser EI, Moser MB (2006) Path integration and the neural basis of the 'cognitive map'. Nat Rev Neurosci 7:663–678.

Miao C, Cao Q, Ito HT, Yamahachi H, Witter MP, Moser MB, Moser EI (2015) Hippocampal remapping after partial inactivation of the medial entorhinal cortex. Neuron 88:590–603.

Moser EI, Moser M-B, McNaughton BL (2017) Spatial representation in the hippocampal formation: a history. Nat Neurosci 20:1448.

Nagelhus A, Andersson SO, Cogno SG, Moser EI, Moser MB (2023) Object-centered population coding in CA1 of the hippocampus. Neuron.

Nash JF (1950) Equilibrium points in n-person games. PNAS 36:48–49.

Neymotin SA, Lytton WW, Olypher AV, Fenton AA (2011) Measuring the Quality of Neuronal Identification in Ensemble Recordings. J Neurosci 31:16398–16409.

Nieh EH, Schottdorf M, Freeman NW, Low RJ, Lewallen S, Koay SA, Pinto L, Gauthier JL, et al. (2021) Geometry of abstract learned knowledge in the hippocampus. Nature 595:80–84.

Niven JE, Anderson JC, Laughlin SB (2007) Fly photoreceptors demonstrate energy-information trade-offs in neural coding. PLoS Biol 5 e116.

O'Keefe J (1976) Place units in the hippocampus of the freely moving rat. Exp Neurol 51:78–109.

O'Reilly KC, Perica MI, Fenton AA (2019) Synaptic plasticity/dysplasticity, process memory and item memory in rodent models of mental dysfunction. Schizophr Res.

Ódor G (2004) Universality classes in nonequilibrium lattice systems. Rev Mod Phys 76:663–724.

Olypher AV, Klement D, Fenton AA (2006) Cognitive disorganization in hippocampus: a physiological model of the disorganization in psychosis. J Neurosci 26:158–168.

Park EH, Kao H-Y, Jourdi H, van Dijk MT, Carrillo-Segura S, Tunnell KW, Gutierrez J, Wallace EJ, et al. (2023) Phencyclidine Disrupts Neural Coordination and Cognitive Control by Dysregulating Translation. Biological Psychiatry Global Open Science.

Park EH, Keeley S, Savin C, Ranck Jr JB, Fenton AA (2019) How the internally organized direction sense is used to navigate. Neuron 101:1–9.

Pavlowsky A, Wallace E, Fenton AA, Alarcon JM (2017) Persistent modifications of hippocampal synaptic function during remote spatial memory. Neurobiol Learn Mem 138:182–197.

Payne HL, Lynch GF, Aronov D (2021) Neural representations of space in the hippocampus of a food-caching bird. Science 373:343–348.

Pehlevan C, Sompolinsky H (2014) Selectivity and sparseness in randomly connected balanced networks. PLoS One 9 e89992.

Poucet B, Hok V (2017) Remembering goal locations. Curr Opin Behav Sci 17:51–56.

Redish AD, Battaglia FP, Chawla MK, Ekstrom AD, Gerrard JL, Lipa P, Rosenzweig ES, Worley PF, et al. (2001) Independence of firing correlates of anatomically proximate hippocampal pyramidal cells. J Neurosci 21:RC134.

Redish AD, Rosenzweig ES, Bohanick JD, McNaughton BL, Barnes CA (2000) Dynamics of hippocampal ensemble activity realignment: time versus space. J Neurosci 20:9298–9309.

Reynolds CW (1987) Flocks, herds and schools: A distributed behavioral model. SIGGRAPH: Computer Graphics, Anaheim. p. 25–34.

Rosser JB (2003) A nobel prize for asymmetric information: the economic contributions of George Akerlof, Michael Spence and Joseph Stiglitz. Rev Polit Econ 15:3–21.

Rueckemann JW, DiMauro AJ, Rangel LM, Han X, Boyden ES, Eichenbaum H (2016) Transient optogenetic inactivation of the medial entorhinal cortex biases the active population of hippocampal neurons. Hippocampus 26:246–260.

Sacktor TC (2011) How does PKMzeta maintain long-term memory? Nat Rev Neurosci 12:9–15.

Samsonovich A, McNaughton BL (1997) Path integration and cognitive mapping in a continuous attractor neural network model. J Neurosci 17:5900–5920.

Sargolini F, Fyhn M, Hafting T, McNaughton BL, Witter MP, Moser MB, Moser EI (2006) Conjunctive representation of position, direction, and velocity in entorhinal cortex. Science 312:758–762.

Savelli F, Yoganarasimha D, Knierim JJ (2008) Influence of boundary removal on the spatial representations of the medial entorhinal cortex. Hippocampus 18:1270–1282.

Schlesiger MI, Boublil BL, Hales JB, Leutgeb JK, Leutgeb S (2018) Hippocampal global remapping can occur without input from the medial entorhinal cortex. Cell Rep 22:3152–3159.
Schneidman E, Berry 2nd MJ, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. Nature 440:1007–1012.
Sheintuch L, Geva N, Baumer H, Rechavi Y, Rubin A, Ziv Y (2020) Multiple maps of the same spatial context can stably coexist in the mouse hippocampus. Curr Biol 30:1467–1476. e1466.
Smith JM (2000) The concept of information in biology. Philos Sci 67:177–194.
Sobel J (2007) Signaling games. In: UCSD (Ed.).
Solstad T, Boccara CN, Kropff E, Moser MB, Moser EI (2008) Representation of geometric borders in the entorhinal cortex. Science 322:1865–1868.
Solstad T, Moser EI, Einevoll GT (2006) From grid cells to place cells: a mathematical model. Hippocampus 16:1026–1031.
Spence M (1973) Job Market Signaling*. Q J Econ 87:355–374.
Talbot ZN, Sparks FT, Dvorak D, Curran BM, Alarcon JM, Fenton AA (2018) Normal CA1 place fields but discoordinated network discharge in a Fmr1-null mouse model of fragile X syndrome. Neuron 97:684–697.
Tian D, Dmitrieva RI, Doris PA, Crary JF, Sondhi R, Sacktor TC, Bergold PJ (2008) Protein kinase M zeta regulation of Na/K ATPase: a persistent neuroprotective mechanism of ischemic preconditioning in hippocampal slice cultures. Brain Res 1213:127–139.
Thomson SR, Seo SS, Barnes SA, Louros SR, Muscas M, Dando O, Kirby C, Wyllie DJA, et al. (2017), Cell-Type-Specific Translation Profiling Reveals a Novel Strategy for Treating Fragile X Syndrome. Neuron 95:550-563 e555.
Tsokas P, Hsieh C, Yao Y, Lesburgueres E, Wallace EJ, Tcherepanov A, Jothianandan D, Hartley BR, et al. (2016) Compensation for PKMzeta in long-term potentiation and spatial long-term memory in mutant mice. Elife 5 e14846.
Ulanovsky N, Moss CF (2007) Hippocampal cellular and network activity in freely moving echolocating bats. Nat Neurosci 10:224–233.
van Dijk MT, Fenton AA (2018) On how the dentate gyrus contributes to memory discrimination. Neuron 98:832–845.
Wesierska M, Dockery C, Fenton AA (2005) Beyond memory, navigation, and inhibition: behavioral evidence for hippocampus-dependent cognitive coordination in the rat. J Neurosci 25:2413–2419.
Wills TJ, Lever C, Cacucci F, Burgess N, O'Keefe J (2005) Attractor dynamics in the hippocampal representation of the local environment. Science 308:873–876.
Witter MP (2006) Connections of the subiculum of the rat: topography in relation to columnar and laminar organization. Behav Brain Res 174:251–264.
Witter MP (2007) The perforant path: projections from the entorhinal cortex to the dentate gyrus. Prog Brain Res 163:43–61.
Yartsev Michael M, Ulanovsky N (2013) Representation of three-dimensional space in the hippocampus of flying bats. Science 340:367–372.
Yoon K, Buice MA, Barry C, Hayman R, Burgess N, Fiete IR (2013) Specific evidence of low-dimensional continuous attractor dynamics in grid cells. Nat Neurosci 16:1077–1084.
Zhenrui L, Darian H, Satoshi T, Ivan S, Attila L (2022) An inhibitory plasticity mechanism for world structure inference by hippocampal replay. bioRxiv. 2022.2011.2002.514897.

APPENDIX 1

Methods

These procedures were previously described in detail (Park et al., 2019), and are here provided in brief to complement the data reported in Fig. 5.

Subjects

We used 14 adult male Long-Evans hooded rats weighing 300–400 g (Taconic Farms, NY). All experimental procedures were approved by NYU's Institutional Animal Care and Use Committee. Rats were handled by the experimenter for 5 days (5 min/day) before surgery under pentobarbital (50 mg/kg, i.p.) anesthesia. The surgery was to implant custom microdrives on the skull that could micro-position electrodes at the recording sites in the brain. Two-four independently movable tetrode-configured electrode bundles were targeted to the hippocampus (relative to Bregma AP 3.8, ML −2.5, DV 1.9) and the medial entorhinal cortex (relative to the from sinus AP 0.5, ML, 4.5, DV, 2.0) of seven rats. Eight independently movable tetrode electrodes were targeted to the medial entorhinal cortex of seven rats. The animals were allowed at least a week to recover before behavioral training began.

Behavior

The rats were habituated to the stainless-steel disk arena (diameter 1.2 m). They were encouraged to continuously forage for 20-mg sugar pellets (Bio-Serv, NJ) that were randomly dispensed to random locations by a computer-controlled overhead feeder. Each of the 5 days of the habituation phase, the arena was stable for 30 min and rotating at 1 rpm for 30 min. Active place avoidance training followed to avoid an annulus-sector that for some rats was 30° and for other rats was 45°. The sector extended from the edge of the arena toward an annulus at either 50% or 40% of the radius, respectively. The rats were trained to avoid a mild < 0.4 mA foot shock if they entered a shock zone that was defined in a fixed location of the room and a coincident fixed location on the arena surface. When the arena was stable the room-defined and arena-defined shock zones were physically identical, but when the arena rotated, they were dissociated such that the stationary room location of the shock zone remained fixed in the room (but not on the rotating arena), and the rotating arena location of shock remained fixed on the arena (but not in the stationary room). The rats received ten daily training sessions consisting of the triad stable1-rotating-stable2, each lasting 30 min, where the arena was stable or rotating and the two stable sessions were identical with the area in an identical orientation. Using two infrared LEDs mounted to the recording electronics on the rat's head, the rat's position and head direction were tracked 30 times a second using an overhead video camera and software (Tracker, Bio-Signal Group, Acton, MA). Position and head direction were tracked in both the stationary spatial frame of the room and the rotating spatial frame of the arena by referencing an infrared LED that was attached to the arena.

Electrophysiology

The electrodes were advanced into the CA1 and MEC regions until action potentials could be recorded. The signals were filtered between 300 Hz and 7 kHz, amplified up to 7000 times and digitized at 48 kHz using commercial hardware and software (dacqUSB, Axona Ltd, St Albans, UK). In this way, ensembles of action potentials were recorded from dorsal hippocampus and dorsomedial entorhinal cortex while the animals performed the active place avoidance task in the stable and rotating conditions. A total of 1465 single units were recorded from MEC, and 756 single units were recorded from CA1. Single unit isolation quality was assessed by computing IsoI_{BG} and IsoI_{NN}, the Isolation Information measures (Neymotin et al., 2011). Only the single units that had values greater than 3.5 bits were considered sufficiently well-isolated for these studies (MEC 810, CA1 414).

Data analysis

Custom C/C + + and MATLAB software was used to compute all outcome measures, with details published, and code publicly available (Park et al., 2019). Functional cell classes were determined by statistical criteria after computing firing-rate tuning as a function of place and head-direction determined by dividing the number of action potentials that a single unit discharged while the rat was in each 2.13-cm square positional bin, or 10° directional bin divided by the total time the rat was detected in that bin. MEC head-direction cells were classified as single units with long duration (> 350 us) action potential and direction-independent firing rate below 10 AP/s. CA1 place cells were classified as single units with long duration action potentials (> 350 us), position-independent firing rate < 5 AP/s, and spatial coherence (> 0.4). The network stability was estimated by computing the n(n-1)/2 pair-wise spike train correlations between all simultaneously recorded pairs of n cells. The spike counts were computed in fixed duration 250 ms time bins for each spike train and Kendall's correlation (τ) was calculated between the spike counts of all pairs of simultaneously recorded cells within MEC, CA1, and between MEC and HPC. The stability of the network state across the Stable1 vs Stable2 and across the Stable1 vs. Rotating conditions was investigated by comparing the τ values for each cell pair in the two conditions, and this was quantified by computing the Pearson correlation across the pairs of τ. The coefficient of determination (r^2) is used to estimate the variance in the τ values in one condition that is explained by the variance in the other condition, and thus the stability of the network discharge. Statistical comparisons between distributions of correlation were performed after transforming r values to z using Fisher's transform.

(Received 17 January 2023, Accepted 6 August 2023)
(Available online 15 August 2023)