PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Research

Cite this article: Schapiro AC, Turk-Browne NB, Botvinick MM, Norman KA. 2017 Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Phil. Trans. R. Soc. B* **372**: 20160049.
http://dx.doi.org/10.1098/rstb.2016.0049

Accepted: 11 August 2016

One contribution of 13 to a theme issue 'New frontiers for statistical learning in the cognitive sciences'.

Subject Areas:
cognition, behaviour, neuroscience

Keywords:
medial temporal lobe, temporal regularities, neural network model, community structure, associative inference

Author for correspondence:
Anna C. Schapiro
e-mail: aschapir@bidmc.harvard.edu

†Present address: Department of Psychiatry, Beth Israel Deaconess Medical Center, 330 Brookline Ave, Boston, MA 02115, USA.

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.3512469.

THE ROYAL SOCIETY PUBLISHING

Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning

Anna C. Schapiro[1,2,†], Nicholas B. Turk-Browne[1], Matthew M. Botvinick[3] and Kenneth A. Norman[1]

[1]Princeton Neuroscience Institute and Department of Psychology, Princeton, NJ 08544, USA
[2]Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02215, USA
[3]Google DeepMind, London EC4A 3TW, UK

ACS, 0000-0001-8086-0331

A growing literature suggests that the hippocampus is critical for the rapid extraction of regularities from the environment. Although this fits with the known role of the hippocampus in rapid learning, it seems at odds with the idea that the hippocampus specializes in memorizing individual episodes. In particular, the Complementary Learning Systems theory argues that there is a computational trade-off between learning the specifics of individual experiences and regularities that hold across those experiences. We asked whether it is possible for the hippocampus to handle both statistical learning and memorization of individual episodes. We exposed a neural network model that instantiates known properties of hippocampal projections and subfields to sequences of items with temporal regularities. We found that the monosynaptic pathway—the pathway connecting entorhinal cortex directly to region CA1—was able to support statistical learning, while the trisynaptic pathway—connecting entorhinal cortex to CA1 through dentate gyrus and CA3—learned individual episodes, with apparent representations of regularities resulting from associative reactivation through recurrence. Thus, in paradigms involving rapid learning, the computational trade-off between learning episodes and regularities may be handled by separate anatomical pathways within the hippocampus itself.

This article is part of the themed issue 'New frontiers for statistical learning in the cognitive sciences'.

1. Introduction

The Complementary Learning Systems (CLS) theory [1] provides a powerful computational framework for understanding the distinct roles that the hippocampus and cortex play in representing memories. It demonstrates the fundamental trade-off between memorizing the specifics of individual experiences (e.g. where I parked my car today), which benefits from separate representations for each experience, and extracting regularities across those experiences (e.g. where in the parking lot spaces tend to be open), which benefits from overlapping representations.

To solve this trade-off, CLS posits that the brain recruits different systems: the hippocampus uses a high learning rate and sparse, relatively non-overlapping (*pattern-separated*) representations to quickly store memory traces for each recent experience without interference from other similar recent experiences. The hippocampus then slowly teaches these experiences to cortical areas during offline periods, such as sleep. The cortex has a slow learning rate and overlapping representations, which allow it to extract and represent regularities across experiences—over days, months, and years.

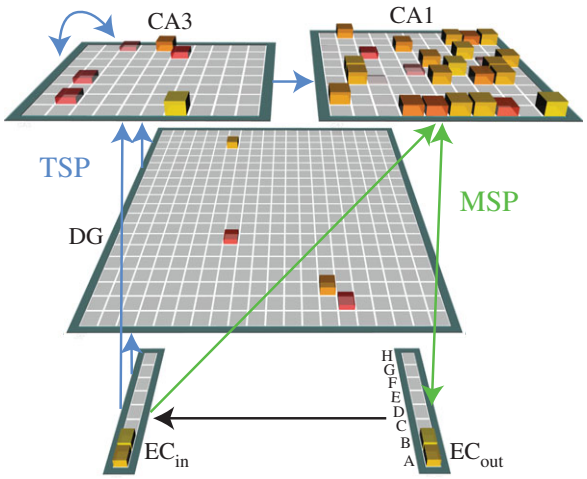

Figure 1. Model architecture. EC_{in} serves as input and EC_{out} as output for the network. The network is trained to reproduce the pattern of activity in EC_{in} on EC_{out}. Three hidden layers—DG, CA3, and CA1—learn representations to support this mapping, with activity flow governed by the projections indicated by the arrows. Blue arrows make up the TSP and green arrows make up the MSP. This snapshot shows network activity during pair structure learning, where pair *AB* is presented to the network and successfully reproduced in EC_{out}. The height and yellowness of a unit both index its activity level.

Although CLS has been successful in accounting for many empirical findings [2], it has not been used to address an important type of learning: apart from memorizing individual experiences rapidly and learning regularities across those experiences over long periods of time, we can also learn regularities rapidly—over minutes or hours [3]. Critically, there have been several empirical demonstrations that the hippocampus is involved in, and even necessary for, such rapid statistical learning [4–11].

These findings pose a challenge for CLS and, more generally, for the traditional view that the hippocampus solely supports memory for distinct episodes. In neural network models, regularities are most effectively extracted via overlapping representations, which facilitate automatic generalization between stimuli based on this shared neural substrate [1]. However, the hippocampus in CLS does not employ overlapping representations—to the contrary, it specializes in minimizing such overlap to prevent interference [1,12]. Thus, although the learning rate of the hippocampus is well suited to the timescale of rapid statistical learning, the mechanism by which it can represent regularities is unclear.

To address this puzzle, we investigate a potential role for the heterogeneous properties across different subfields and pathways of the hippocampus. We ran simulations of statistical learning using a recent version of a hippocampal neural network model that instantiates the episodic learning component of CLS [13]. The model explains how known anatomical pathways in the hippocampus and subfield properties might together support episodic memory. It contains hippocampal subfields dentate gyrus (DG), cornu ammonis 3 (CA3), and CA1, as hidden layers that learn to map input provided by superficial layers of entorhinal cortex (EC_{in}) to output through deep layers of EC (EC_{out}; figure 1). Layers in the model are composed of *units*, which simulate a rate code of the spiking activity of neurons or small populations of neurons. Units in different layers can be linked with weighted connections, which represent the efficacy of the synapse(s) between the neurons and determine how activity spreads from one unit to another.

There are two main pathways connecting these subfields in the hippocampus, and correspondingly in the model: the trisynaptic pathway (TSP), $EC_{in} \rightarrow DG \rightarrow CA3 \rightarrow CA1$, and the monosynaptic pathway (MSP), $EC \leftrightarrow CA1$. The sets of connection weights, or *projections*, between layers within the TSP are sparse (a small subset of units are connected between layers), and CA3 and especially DG have high levels of within-layer inhibition. High connection sparsity and high inhibition result in few units being active at any time in DG and CA3 (figure 1) and allow the layers to avoid interference by forming separated, conjunctive representations of incoming patterns, even when the patterns are highly similar. The projections within the TSP also have a very high learning rate. These properties together make the TSP the engine for rapid encoding and pattern separation of episodic memories in the hippocampus. The TSP, with the help of strong recurrent connections within CA3, also serves to retrieve previously memorized patterns from partial cues (*pattern completion*).

By contrast, the projections within the MSP are not as sparse, and the inhibition is lower, allowing more units to be active at any time in CA1 (figure 1). The MSP projections also have a relatively lower learning rate. The higher level of pattern overlap (due to more units being active for each input) and the lower learning rate make CA1 more cortex-like. Indeed, the region has acted in extant models as a translator between sparse representations in the TSP and overlapping representations in EC. The role of the MSP has simply been to help the hippocampus communicate with cortex.

To investigate how these different projections and subfields might support rapid statistical learning, we chose three learning paradigms for simulation that require extracting regularities on the timescale of minutes to hours. All have been linked to the hippocampus, with representations of associated stimuli coming to be represented more similarly [5,6,14]. The first, *pair structure*, is an example of the classic statistical learning paradigm [6,15–17], in which a continuous sequence of items is presented to participants during passive viewing or a cover task. Unbeknownst to them, the sequence contains embedded regularities in terms of which items tend to follow each other. Pair or triplet regularities (pairs in our simulation) are each composed of a fixed set of arbitrarily chosen items that repeatedly occur in the same order, while the order of the pairs/triplets is random. Because items are presented continuously, one at a time, and there are no timing or other cues as to the event boundaries, this task requires integrating statistics over time in order to uncover the regularities—namely, the higher transition probability within versus between pairs. An additional target for our simulations is the finding that infants show learning in this paradigm despite underdeveloped TSPs [17,18].

The second paradigm, *community structure* [5,19], again involves a continuous sequence of items with hidden groupings, but in contrast, the sequence is generated such that transition probabilities between adjacent items are uninformative about these groupings, and instead higher level sensitivity to associations *shared* across items is required to learn the groupings. This paradigm allows us to explore how higher level learning might occur in the model, and in the hippocampus.

Finally, there is a class of learning paradigms, including *transitive inference*, *acquired equivalence*, and *associative inference*, that also probe memory for indirect associations

(i.e. associations not based on directly observed item pairings, but rather on the fact that certain pairs share associates). Although these paradigms are not considered conventional statistical learning paradigms, in that there is no need for segmentation from otherwise undifferentiated input, they share the requirement for rapid integration across experiences over time and thus may tap into the same learning mechanisms as the pair and community structure paradigms above [20]. To explore this possibility, we include a simulation of associative inference.

When confronting the hippocampal model with these different kinds of structured input, we found that—in addition to its role translating between the TSP and EC—the MSP took on a new role: it learned and represented the regularities. Both representational similarity and recurrent activity dynamics contributed, making contact with a recent computational model of how recurrent activity in the hippocampus can support generalization [21]. Combined with the role of the TSP in rapid learning of the specifics of individual experiences, these findings reconcile the trade-off between episodic memory and statistical learning by suggesting that the hippocampus itself contains complementary learning systems.

2. Material and methods

(a) Model architecture

We built on a neural network model of the hippocampus that has been successful in accounting for episodic memory phenomena, and incorporates known projections and properties of hippocampal subfields [12,13,22]. We used a recent implementation [13] in the Emergent simulation environment [23], v. 7.0.1. Our project was modified from hip.proj [24], as it was implemented for version 6 of Emergent. See the electronic supplementary material, tables S1 and S2, for parameter details.

(i) Activity dynamics

The model is composed of units with activity levels ranging from 0 to 1. A unit's activity is proportional to the activity of all units connected to it, weighted by the value of each of the connection weights between them. A unit's activity is also modulated by local inhibition between units within a layer. The inhibition corresponds to the activity of inhibitory interneurons, implemented in these simulations using a set-point inhibitory current with a k-winners-take-all dynamic. (See [24,25] for details and equations.)

(ii) Input/output

Superficial layers of EC (EC_{in}) provide input to the hippocampus and deep layers of EC (EC_{out}) provide output. EC_{in} and EC_{out} use orthonormal representations in these simulations—each item in the paradigm was represented by activation of one unit (with the number of units in EC_{in} and EC_{out} varying across paradigms). There is also a separate Input layer, not shown, with the same number of units as EC_{in} and one-to-one connections to EC_{in}. Input was clamped in this layer so as to allow EC_{in} to also receive input from EC_{out}, completing the 'big loop' of the model. To clamp input, the units in the Input layer corresponding to the current and previous stimulus were forced to maintain high activity, while all other units were forced to have no activity. Activity travelled from the Input layer to the stimulus representations in EC_{in}, then to the rest of the network. EC_{in} and EC_{out} each had inhibition set so that two units could be active at a time ($k = 2$), unless otherwise noted.

(iii) Trisynaptic pathway

EC_{in} projects to DG and CA3 in the TSP. These projections are sparse, reflecting known physiology. Each DG and CA3 unit receives input from 25% of the EC_{in} layer, and the 'mossy fibre' projection from DG to CA3 is even sparser (5%). DG and CA3 additionally have high levels of within-layer inhibition. CA3 also has a fully connected (every unit to every other) projection to itself, which helps bind pieces of a representation to one another and retrieve a full pattern from a partial cue. CA3 then has a fully connected projection to CA1, completing the TSP.

(iv) Monosynaptic pathway

There are fully connected projections in the MSP from EC_{in} to CA1, CA1 to EC_{out}, and EC_{out} to CA1. CA1 has much less local inhibition than DG and CA3. See the electronic supplementary material for discussion of the removal of the MSP 'slots' used in previous versions of the model.

(v) Network initialization

For each simulation, we ran 500 networks. Each network corresponds to a particular randomized configuration of the sparse projections in the TSP and to randomly re-initialized weights throughout the network. All analyses were done within each of the 500 networks, and results were then averaged across networks. These randomly re-initialized networks were treated as random effects in statistical tests.

(b) Learning

The model is trained to adjust connection weights between units such that it can duplicate the patterns presented to EC_{in} on EC_{out}. The learning in the version of the model used here [13] is different from older versions in two important ways. First, whereas previous versions used only Hebbian learning, this version uses a combination of error-driven and Hebbian learning [12,13,25]. The error-driven component uses Contrastive Hebbian Learning [26], which adjusts connection weights such that activity during a 'minus phase' becomes more similar to activity during a 'plus phase.' In the minus phase, the model uses its existing connections to try to produce the correct output given the input, and in the plus phase, the model is directly shown the correct output. The difference between the minus and plus phase activity patterns is the *error*. Adjusting weights so that minus phase activity looks more like plus phase activity results in the model being more likely to produce the correct output in the future. Incorporating this error-driven learning into the model results in increased memory capacity [13]. The second update to the learning algorithm is that previous versions pre-trained the MSP weights, whereas this version trains the MSP online, at the same time as the TSP, which also results in better memory capacity [13].

The learning procedure used here is based, as in other models [27], on empirical findings of differences in projection strengths between subfields at different phases of the hippocampal theta oscillation [28]. At the trough of the theta cycle, as measured at the hippocampal fissure, EC has a stronger influence on CA1, whereas at the peak, CA3 has a stronger influence on CA1. The model instantiates these two phases of theta as two minus phases on each trial. In one, EC_{in} projects strongly to CA1, and CA3 → CA1 is inhibited; this corresponds to a discrete sample of the theta trough. In the other, CA3 projects strongly to CA1, and EC_{in} → CA1 is inhibited; this corresponds to a discrete sample of the theta peak. These phases have been proposed to correspond to encoding- and retrieval-like states: a strong influence of EC_{in} on CA1—a state where external input directly influences CA1—is considered more akin to encoding, while a strong influence of CA3 on CA1 is more akin to retrieval [27].

Activity during each of the two minus phases in a trial is contrasted with activity during a plus phase, in which the target pattern is directly clamped on EC_{out}. Weights are changed after each trial such that patterns of unit coactivity during each minus phase are shifted more towards those of the plus phase.

Modification of the model's internal representations to better align with the observed environment is a general property of error-driven learning algorithms, but in the case of the hippocampus may be related to the idea that the region carries out match/mismatch computations [29,30]. The learning rate in the TSP is set to be 10× higher than in the MSP. See Ketz *et al.* [13] and the electronic supplementary material, table S2 for more details. Note, however, that we use the version of the model implemented in O'Reilly *et al.* [24], which uses Contrastive Hebbian Learning for all learnable projections, as opposed to pure Hebbian learning used for some TSP projections in Ketz *et al.* [13] (though the results do not depend on this difference).

(c) Stimulus presentation

To simulate sequential item presentation, we presented items to EC_{in} using a moving window that encompassed the current and previous items. This approach is based on findings that the representation of a previous stimulus persists over delays in EC [31,32]. Lesions to parahippocampal regions but not to hippocampus proper impair delayed non-match to sample performance [33–35], suggesting that these regions may be more important than the hippocampus in maintaining simple traces of past stimuli.

We included temporal asymmetry, presenting the current item with full activity (clamped value = 1) and the previous item with decayed activity (0.9). During training, one *trial* (lasting 100 processing cycles) consisted of the presentation of two such items for two minus phases and one plus phase.

(d) Testing and analyses

We tested the network with trials in which each stimulus was presented by itself as the input, without the trailing activity of the preceding item and without a target pattern. On each trial, we clamped the activity of the Input layer unit for the stimulus to 1 and recorded the activity level of each unit throughout the network after 20 cycles of processing (initial response), and after the network had fully settled (we used 80 cycles, though 40–50 cycles are usually enough). We chose 20 cycles for the initial response because at that point activity has spread throughout the network, including to EC_{out}, but not from EC_{out} to EC_{in} (preventing big-loop recurrence). For some simulations, we also looked at the activity evoked by pairs of items, in which case input units for both items were clamped to 1.

During test, network representations were assessed in a neutral state, where the projections between layers were scaled to values in between the encoding-like and retrieval-like states described in the *Learning* section. No connection weights were changed during testing. Networks were tested before training (epoch 0) and after every *epoch* of training. An epoch is a grouping of multiple trials, with the number of trials per epoch specified below for each simulation.

(i) Output analyses

We assessed the probability, after settling, of activating a particular item in EC_{out} above 0.5 given presentation of a particular item in EC_{in}.

(ii) Pattern analyses

We recorded the pattern of unit activity evoked by each test item, in each hidden layer, for the initial and settled response. We calculated Pearson correlations between the patterns evoked by different items.

(e) Lesions

To help assess the contribution of particular pathways to network behaviour, we simulated lesions. To lesion the TSP, we set the strength of the following projections to 0: $EC_{in} \rightarrow DG$, $EC_{in} \rightarrow CA3$, $DG \rightarrow CA3$, $CA3 \rightarrow CA3$, and $CA3 \rightarrow CA1$ (all blue projections in figure 1). To lesion the MSP, we set the strength of $EC_{in} \rightarrow CA1$ to 0. We did not alter the $CA1 \rightarrow EC_{out}$ projection, though it could be considered part of the MSP, because this projection is required for producing output in the model. It is important to note that although TSP lesions can reveal the independent functions of the MSP, a MSP lesion does not cleanly reveal the contribution of the TSP, as the MSP always serves to facilitate communication between the TSP and EC. Pathways were lesioned during both training and testing. Other reported lesions were similarly implemented by setting the strength of specified projections to 0.

3. Results

(a) Learning episodes versus regularities

We tested whether a model of the hippocampus designed to simulate episodic memory [12,13] can pick up on statistics in continuous sequences. In prior simulations using this model, the episodes to be learned were clearly demarcated for the model; for example, the model was used to simulate experiments in which subjects were shown word pairs and asked to memorize that they go together [12]. By contrast, in statistical learning experiments, there is a continuous sequence of stimuli with no demarcations of event boundaries. The only way to detect the regularities is to track statistics across experiences.

We exposed the model to sequences containing embedded pairs. There were eight items (*A*–*H*) grouped into four pairs (*AB*, *CD*, *EF*, *GH*). Items within a pair always occurred in a fixed order but the sequence of pairs was random. Specifically, the second item in a pair could transition to the first item in one of the three other pairs. Back-to-back repetitions of a pair were excluded, since this is a common constraint in statistical learning experiments and because allowing repetitions would dilute the temporal asymmetry (both *AB* and *BA* would be exposed). There was a moving window of two stimuli presented at a time. After *AB*, for example, *BC*, *BE*, or *BG* followed with equal probability; if *BC* was chosen, the next input would be *CD*. To detect regularities, the model had to be sensitive to the fact that, over time, pairs (e.g. *AB*) occurred more often than sets of two items spanning pairs (e.g. *BC*). To contrast this learning challenge with a more 'episodic' situation with demarcated events, we also ran simulations where pairs were presented in interleaved order, with no transitions between pairs. In other words, *AB*, *CD*, *EF*, and *GH* all appeared but never *BC* or *FG*, for example. Note that we will use *A* and *B* to refer to the first member and second member of a pair, respectively. All results are averaged across the four pairs.

Different representations emerged in the model for sequences that did versus did not require sensitivity to statistics across trials. For the latter, 'episodic' sequences (with demarcated pairs), the network quickly learned to activate both members of a pair in EC_{out} (*A* and *B*), when presented with either pairmate individually (*A* or *B*; figure 2*c*). The hidden layers learned representations that allowed the model to perform this mapping between EC_{in} and EC_{out} (figure 2*a*). In particular, CA3 and DG in the TSP rapidly memorized conjunctive representations of each pair (i.e. memorized the distinct pattern of activity evoked by presentation of both pair members at once). Because each item had only been viewed in one pair, this memorization caused each pairmate, on a given test trial, to immediately pattern complete to this

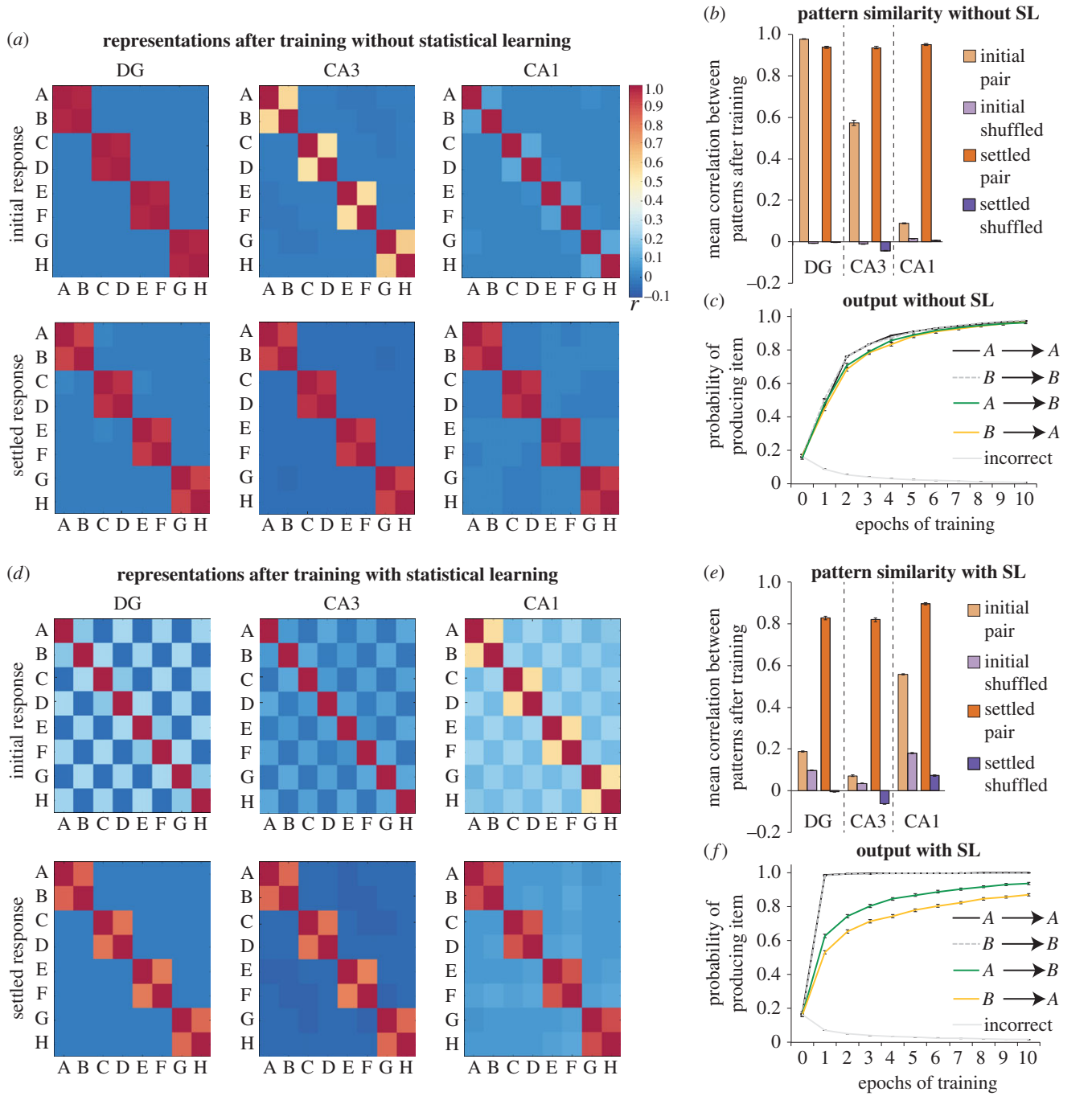

Figure 2. Pair structure. (*a*) Average representational similarity across networks in each of the three hidden layers of the model, after training on episodic sequences that did not require statistical learning (SL). In the heatmaps, each of the eight test items appears in the rows and columns, the diagonals correspond to patterns correlated with themselves, and the off-diagonals are symmetric. (*b*) Average representational similarity by pair type, for the initial and settled response. 'Shuffled' pairs are items paired with all other items that were not the trained pairmate (e.g. *AC*, *DA*), including both viewed pairings (e.g. *DA*) and unviewed pairings (e.g. *AC*). (*c*) Average probability of activating a particular item on the output given a particular item on the input, over training. For example, $A \rightarrow B$ is the probability of activating the second member of a pair above threshold given the first. 'Incorrect' is the probability of producing an item that is not the current item or its pairmate. Each input was presented once per epoch in permuted order. (*d*–*f*) Same as above, for sequences that required SL. Each pair was presented approximately five times per epoch (with 80 total inputs per epoch). For all subplots, values are means across 500 random network initializations. Error bars denote ± 1 s.e.m. across network initializations. Some error bars are too small to be visible.

conjunctive representation. CA3's initial pair similarity was somewhat lower than DG's because the two sets of sparse projections—$EC_{in} \rightarrow DG$, $DG \rightarrow CA3$—make it more difficult for CA3 to integrate across paired items.

Initial similarity in CA1 was much weaker after training (figure 2*b*) simply because CA1 learns more slowly (CA1 initial pattern similarity becomes higher after longer training, not shown), but this slower learning rate does not detract from CA1's ability to help communicate information in CA3 to EC_{out}. After the pattern is completed in EC_{out} and information travels to EC_{in}, the similarity structure is then strongly apparent in the settled response in all hidden layers.

What happens when the events to be memorized are not demarcated and must be learned over time from temporal statistics? The model was still able to learn the pairs, activating *B* on EC_{out} given *A* on EC_{in} at test, and vice versa (figure 2*f*), despite each of those items also being exposed with three other items. Unlike the previous simulation, there was a

striking lack of pair-related similarity in the initial response in DG and CA3 after training (figure 2*d*). Pair similarity was slightly higher than shuffled pair similarity (figure 2*e*) but not for the subset of shuffled pairs that appeared as across-pair transitions (e.g. *DA*; electronic supplementary material, figure S1).

The checkerboard structure in DG and CA3 reflects the fact that both *AB* and *DA* occurred in the sequence, but never *AC*. That is, these regions are sensitive to which items have co-occurred, but do not preferentially retrieve an item's most frequently co-occurring mate. DG memorizes every exposed input pattern, and when given a cue like *A* that is part of more than one pattern, activates units involved in all conjunctive representations of exposed patterns that included *A* (including *AB* but also the other exposed patterns *DA*, *FA*, and *HA*). After the network retrieves *B* on EC_{out} (due to the influence of CA1, as described below), *B* in addition to *A* becomes active in EC_{in}, and the DG representation shrinks down to only the units involved in the *AB* conjunctive representation. For this reason, pair structure emerges after the network has settled. The dynamics are similar though not quite as clean in CA3.

In contrast to the previous simulation, CA1 here showed robust similarity for the paired items in its initial (and settled) response. This is because there was enough exposure for CA1 to learn representations with its slower learning rate, and because distributed, overlapping representations are highly sensitive to frequencies of occurrence [36].

These simulations indicate that the TSP and MSP form different representations depending on the learning problem. When sensitivity to statistics is not required, the TSP memorizes each of the presented patterns and can complete them from a partial cue, while CA1 simply acts as a translator between CA3 and EC. When sensitivity to statistics is required, the TSP unhelpfully captures both the pairs and the transitions between them, while the MSP is able to represent the frequencies of item co-occurrence. In both cases, pattern completion expressed in EC_{out} causes the pairmate to arrive on EC_{in}, and the pair's representation activates throughout the network.

(i) Temporal asymmetry

Paired items always appeared in a fixed order *AB* and never in the reverse order *BA* (because back-to-back pair repetitions were not allowed), with the previous item less active than the current item. Thus, learning of the pairs had the potential to be asymmetric in time, with either a forward bias or a backward bias. We found the former: *B* was more likely to be activated in the output layer given *A* as input than *A* was given *B* (figure 2*f*). This results from the higher activity of *B* compared to *A* when *AB* is presented in training, causing more strengthening of the weights connecting to *B*'s output unit than to *A*'s. This is consistent with findings that the hippocampus is involved in prediction rather than retrodiction [6,7,37–39].

(ii) Representational change over time

Though the TSP does not represent regularities in the initial response at the end of training, there is an earlier period where it does represent them weakly, and in fact the effect in this period is stronger in DG than CA1 (electronic supplementary material, figure S1*a*). This occurs because the TSP rapidly learns whatever it is exposed to, and the true pairs *AB* do occur more frequently. However, as the network gains more exposure to all transitions (*BC*, *DA*, etc.), weights max out and the similarity for across-pair transitions fully catches up. As *B* pulls away from *AB* and towards *BC*, the relatively greater similarity for pairmates weakens. CA1's similarity structure does not weaken over time in this way, even with extensive training (electronic supplementary material, figure S1*c*,*d*). To assess the impact of these TSP dynamics on the model's behaviour, we ran simulations with the MSP lesioned. The TSP alone initially supported weak retrieval of the correct pairmate (and showed the same temporal asymmetry as the intact network), but over time, within- and across-pair transitions were retrieved equally (electronic supplementary material, figure S1*b*). Thus, by the end of training, a network with an MSP lesion exhibits virtually no sensitivity to regularities.

(iii) Statistical learning with an undeveloped TSP

Infants are good at statistical learning despite having immature hippocampi [16,17], and our model suggests that this may be because the MSP develops earlier than the TSP [18,40]. To test the boundary case of infants only having access to the MSP (though they may have some limited use of the TSP), we ran additional simulations with the TSP lesioned. We found that learning and representations were essentially unchanged (electronic supplementary material, figure S2). In fact, pairmate retrieval was slightly better without versus with the TSP (at the end of training, mean $A \rightarrow B$ retrieval: 0.98 versus 0.94, respectively, $t_{998} = 6.66$, $p < 0.001$; mean $B \rightarrow A$ retrieval: 0.90 versus 0.87, $t_{998} = 2.49$, $p = 0.01$). This stands in contrast to the very detrimental effects of an MSP lesion (electronic supplementary material, figure S1*b*), revealing that the MSP is necessary and sufficient for statistical learning and providing a possible explanation for intact rapid statistical learning in infants despite protracted hippocampal development.

(b) Higher level learning

The structure of the sequences considered so far was simple—it could be learned by tracking the strength of transition probabilities between adjacent items or the joint frequencies of pairs. To examine how the MSP handles higher level, more complex statistics, we simulated learning of a 'community structure' sequence that cannot be parsed based on transition probability or joint frequency [5,19]. The sequence was generated via a random walk on a graph (figure 3*b*) with three densely interconnected 'communities' of nodes [41]. The walk tends to stay in a community for a while before transitioning to the next, but any individual node has an equal probability of transitioning to exactly four other nodes. We constrained the sequences such that the observed joint frequencies were exactly equated in each epoch (as opposed to equated on average). Thus, transition probabilities did not provide information about the location of community boundaries. To learn these boundaries, it was necessary to pick up on the fact that nodes in the same community were connected to overlapping sets of other nodes, whereas nodes from different communities were not.

The only change to the architecture and parameters for this simulation was the addition of seven units to EC_{in} and EC_{out} to accommodate the 15 nodes in the graph (versus the eight paired items). We exposed the model to sequences in the same way as for the pair structure paradigm, with two items presented at a time in a moving window and the previous

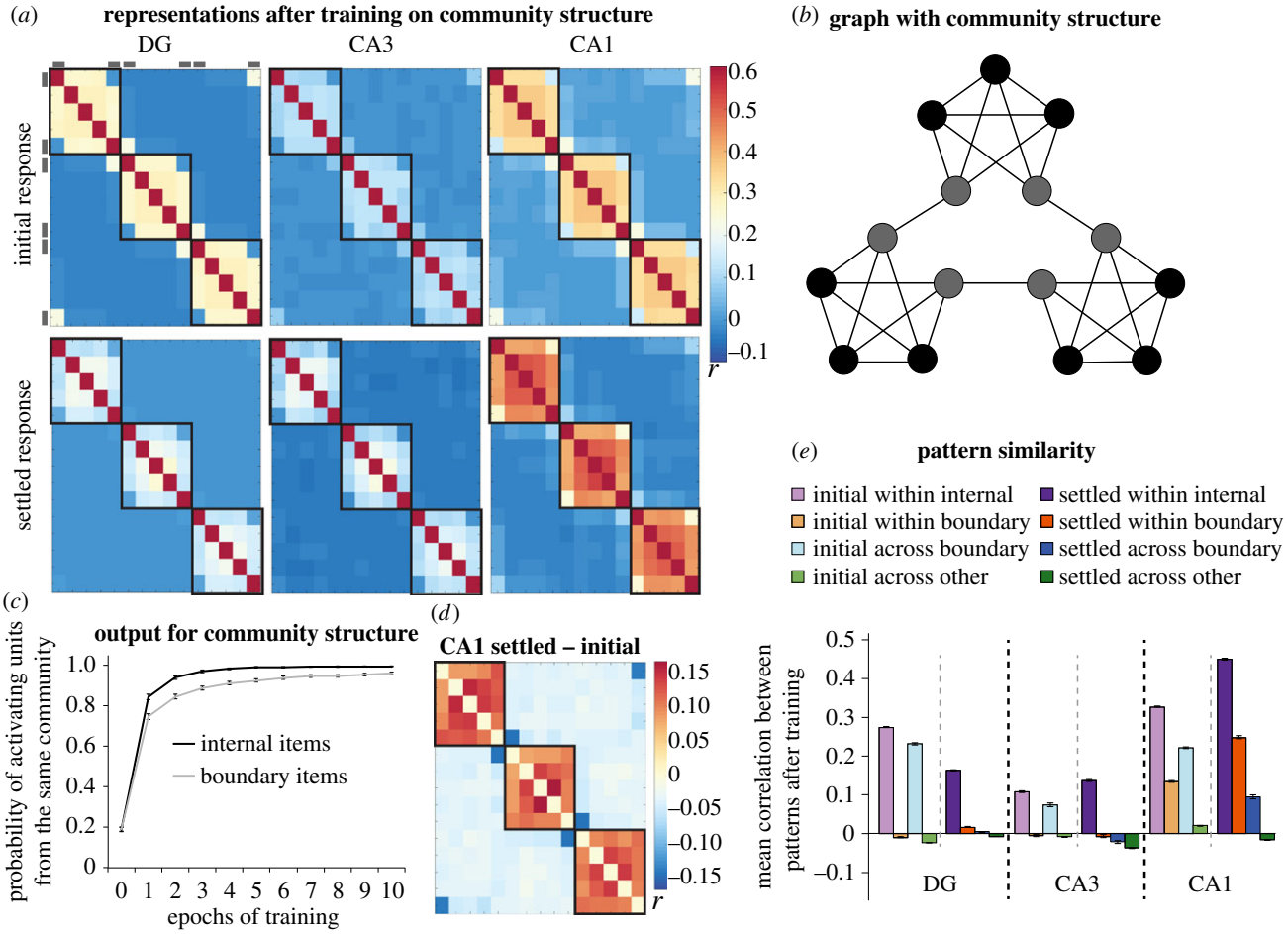

Figure 3. Community structure. (*a*) Average representational similarity after training, with items arranged by community (black boxes; grey bars mark boundary nodes). (*b*) Graph with three communities of nodes. Each node on the graph represents a particular item, and the edges indicate which transitions were allowed (bidirectionally). (*c*) Average probability of activating units from the same community given an internal item (black node) or boundary item (grey node) as input, over the course of training. Values are lower than chance (0.33) prior to training due to units not reaching 0.5 threshold level of activity. (*d*) Difference between the settled and initial heatmaps in CA1. (*e*) Average representational similarity between two neighbouring nodes from the same community (within internal), the two boundary nodes from the same community (within boundary), two adjacent boundary nodes from different communities (across boundary), and all other pairs of items from different communities (across other).

item less active than the current item. Note that there was no temporal asymmetry in the stimulus sequences for this simulation, though, as the edges in the graph were bidirectional and thus every pair of connected nodes occurred in both orders. There were 60 inputs per epoch, and 10 total epochs.

The model successfully learned to activate other items from a test item's community (figure 3*c*). Notably, this was still true when restricting analysis to the boundary items: by the end of training, tested boundary items almost never activated the adjacent boundary node in a different community in preference to one of the adjacent nodes in the same community. The network again exhibited different representational similarity in the initial and the settled response (figure 3*a*): DG and CA3 initially represented all exposed pairs but largely ignored higher level structure; by contrast, CA1 was immediately sensitive to higher level structure.

This is best illustrated at community boundaries: two adjacent boundary nodes were directly observed together during training despite being from different communities, whereas two boundary nodes in the same community were never seen together in training and could only be associated *transitively* through other nodes in the community. Thus, high similarity for within-community boundaries means that a region has learned structure beyond basic co-occurrence statistics. This was found in the initial response in CA1 (figure 3*e*), although it also had high similarity for across-community boundaries. DG and CA3 showed low similarity for within-community boundaries and high similarity for across-community boundaries, indicating that these regions had poor sensitivity to the higher level structure.

Consistent with these findings of robust higher level representations in the initial response of the MSP but not TSP, lesioning the TSP did not affect network behaviour (probability of activating a node from the same community given a boundary node = 0.95, versus intact: $t_{998} = 1.74$, $p = 0.083$), whereas lesioning the MSP resulted in very poor behaviour (0.63, versus intact: $t_{998} = 16.6$, $p < 0.001$). Note that this probability being lower than 0.75 indicates that the MSP lesion also degraded the model's ability to learn direct adjacencies, as choosing randomly among a boundary node's neighbours would produce a 0.75 probability.

After allowing big-loop recurrence, the higher level structure became clearer in CA1, and throughout the network. Community structure fully overtook adjacency, with an increase in within-community boundary similarity and a decrease in across-community boundary similarity (figure 3*d*,*e*). To understand why, consider what happens when a boundary item is presented at test: it retrieves a directly linked associate, which is more likely to be an internal node from the same community than the boundary node from the other community, because of

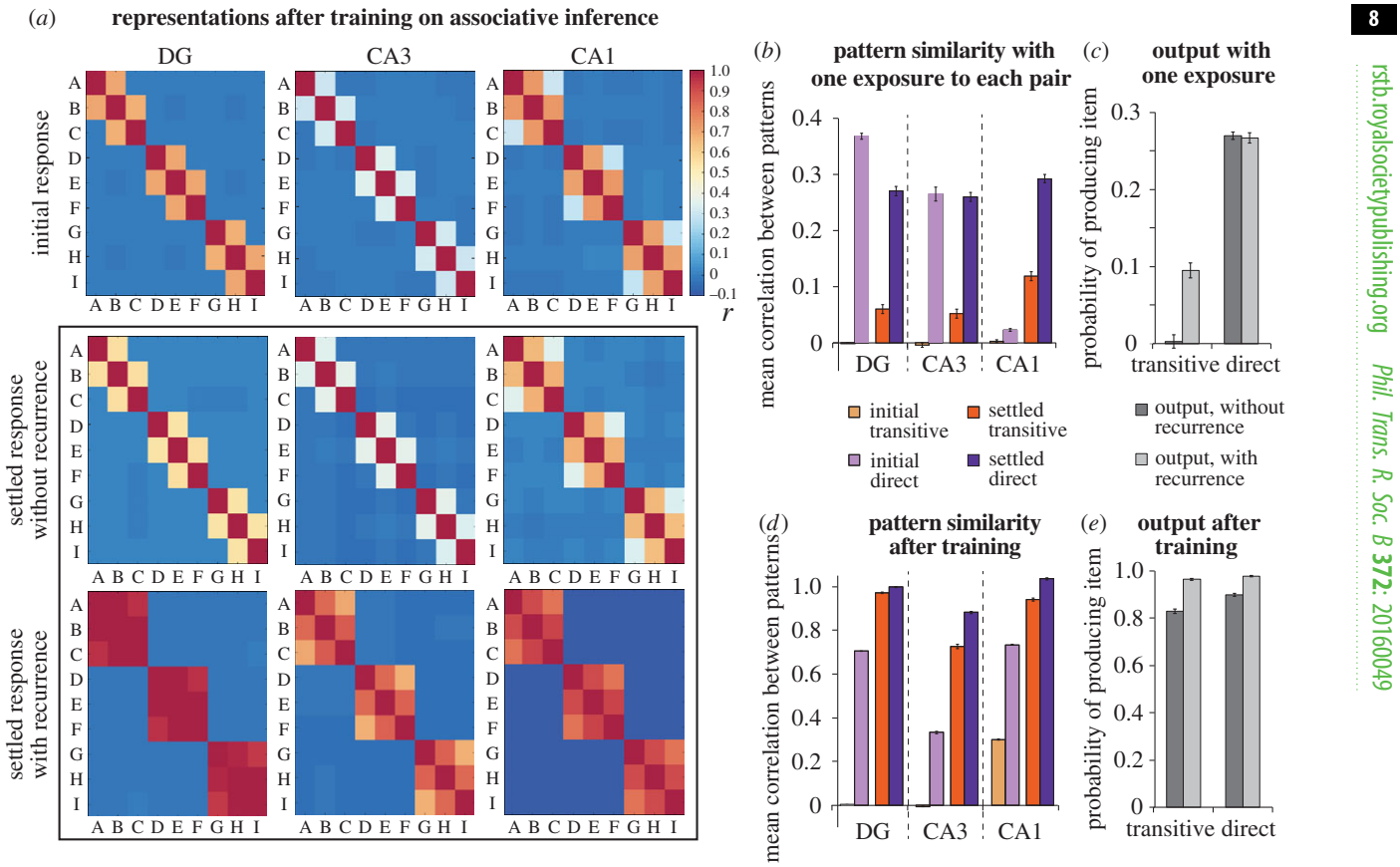

Figure 4. Associative inference. (*a*) Average representational similarity after training, with items arranged by triad. The settled response is shown with and without recurrence allowed. The initial response was very similar for the two variants and is shown with recurrence. (*b*) Similarity structure after one training trial with each of the direct pairs. The pattern correlations between members of direct *AB* pairs and between members of transitive *AC* pairs are shown, subtracting the correlation for shuffled pairs as a baseline. This is shown for the initial and settled response, in both cases in networks with recurrence allowed (though only the settled response is affected by recurrence). (*c*) The probability of producing the direct pairmate (*B* given *A*) and transitive pairmate (*C* given *A*), subtracting the probability of producing other items as a baseline, with and without recurrence. We allowed any above-zero activity in EC_{out} units to count as 'producing the item', simulating the sensitive forced choice test used in associative inference studies [14]. (*d*,*e*) Same as *b*, *c* for fully trained network.

greater within-community internal versus across-community boundary similarity (figure 3*e*). EC_{out} will send activity for that internal node to EC_{in}, which then travels to the hidden layers and further emphasizes the representations of nodes in that community, including the other boundary node in the community that is only associated transitively through other community members. We verified that the difference between the initial and settled representational similarity was indeed attributable to big-loop recurrence (versus recurrence in activity between CA1 and EC_{out}), by lesioning either $EC_{out} \rightarrow EC_{in}$ (needed for big-loop recurrence) or $EC_{out} \rightarrow$ CA1 during testing. Higher level behaviour was unchanged in the $EC_{out} \rightarrow$ CA1 lesioned network but much weaker in the $EC_{out} \rightarrow EC_{in}$ lesioned network.

(c) Associative inference

We next simulated an associative inference paradigm, which requires higher level integration as in the community structure paradigm, but without the continuous presentation of individual items characteristic of temporal statistical learning. We ran these simulations to explore whether the MSP learning mechanisms implicated above in statistical learning can also subserve other related forms of learning. In associative inference, two items *A* and *B* are presented together, and in separate trials, *B* and *C* are presented together. Humans and animals are then able to infer that *A* goes with *C* [20].

We trained the model on three such triads: *ABC*, *DEF*, and *GHI*. We presented each *AB* and *BC* pair 10 times per epoch in permuted order, over 20 epochs. Because the paradigm involves presenting two stimuli at a time, both stimuli were presented with full activity in the input. We made one change to the network parameters in addition to changing the number of units in EC_{in} and EC_{out} to 9: to allow the model the possibility of activating the transitive associate (*C* given *A*) and not just itself and its direct associate (*A* and *B* given *A*), which is its strong tendency with $k = 2$, we lowered inhibition to $k = 3$ in EC_{in} and EC_{out} when testing the network (during training, $k = 2$, as before). This is critical for accounting for transitive behaviour. See the electronic supplementary material, figures S3 and S4 for discussion of this point and simulations showing how $k = 3$ at test affects pair structure and community structure simulations (in short, it enhances higher level structure in the latter and adds a small amount of noise to the former).

The results of these simulations matched the community structure simulations. In the initial response, DG/CA3 represented direct pairs *AB* and *BC*, but not transitive pair *AC*, whereas CA1 had graded similarity structure that reflected transitive relationships (figure 4*a*).

There has been recent debate as to whether this task is solved using recurrence or 'static' representational similarity—overlap in representational space that is not a result of recurrent activity dynamics [20,21,32,42]. To explore this

issue in our model, we contrasted pattern similarity in the settled response for networks with and without recurrence during test. Networks without recurrence had $EC_{out} \rightarrow EC_{in}$ and $EC_{out} \rightarrow CA1$ connections lesioned (only in the test phase), so that representational similarity and behaviour must be caused by the representations in the initial response. The transitive structure was present without recurrence in CA1, but became stronger throughout the network with recurrence (figure 4*a*), indicating that both static representational similarity and recurrence contribute to transitive associations in the model.

Another way to evaluate the role of recurrence is to assess representations and behaviour very early in training, before CA1 has time to develop overlapping representations for transitive associations. After just one exposure to each of the direct pairs, when there was no transitive structure at all in the initial response in CA1 (figure 4*b*), recurrence could support transitive behaviour (figure 4*c*). Indeed, recurrence was *necessary* for this behaviour, as the network without recurrence could produce direct pair but not transitive output after one exposure. The network can use rapidly memorized associations between the direct pairs in the TSP in addition to recurrent dynamics to link from *A* to *C* through *B*. After recurrent activity settles, the transitive associate appears in EC_{in}, allowing all hidden layers to show apparent transitive pattern similarity (figure 4*b*). CA1 notably showed stronger settled similarity here than DG and CA3 because of its lower sparsity.

After full training, CA1 showed transitive similarity in the initial response (figure 4*a*,*d*). This representation was sufficient to support transitive behaviour, as transitive associates were produced in EC_{out} even when recurrence was not allowed (figure 4*e*). Allowing recurrence enhanced the similarity structure throughout the network (figure 4*a*,*d*) and increased transitive behaviour (figure 4*e*).

These simulations suggest that both the TSP and MSP can solve associative inference by the end of training, just using different strategies: the TSP uses recurrence-based dynamics, whereas the MSP uses static representational similarity. Indeed, at the end of training, with a lesioned TSP, the network produces the correct transitive associate with a probability of 0.99 on average (versus 0.01 for incorrect items, $t_{499} = 386.1$, $p < 0.001$). Lesioning the MSP—though a less clean manipulation, as the MSP contributes to TSP communication with EC—still results in production of the correct transitive associate with probability 0.66 on average (versus 0.20 for incorrect items, $t_{499} = 32.4$, $p < 0.001$).

4. Discussion

Avoiding interference between related memories is often critical, but it can be just as important to notice their commonalities. Our simulations suggest that there may be complementary learning systems within the hippocampus that achieve these competing goals—a microcosm of the broader CLS theory of hippocampus and cortex. We found that DG and CA3, forming the TSP to area CA1, represented distinct episodes but failed to learn regularities across episodes. This reflects aggressive pattern separation of similar experiences in these regions, caused by sparse connectivity and high inhibition. By contrast, the direct MSP pathway to CA1 learned regularities across experiences. This is facilitated by overlapping representations in CA1 that result from full connectivity, lower inhibition, and a slower learning rate. The MSP can function even when the TSP is fully lesioned, which might provide an explanation for why infants, who have undeveloped TSPs, are prodigious statistical learners. The overlapping representations in CA1 are akin to 'nodal codings' that have been observed there, with shared features across events represented by overlapping populations of neurons ([43,44], see also [45]). These representations have been proposed to support relational memory and generalization [46], consistent with our account.

The community structure and associative inference simulations demonstrated that the model is also capable of learning transitive associations. For community structure, CA1 learned to represent the two boundary nodes in the same community more similarly despite the fact that they were never experienced together. This was possible because they shared overlapping sets of associates—a principle we previously demonstrated in a simpler neural network model of temporal community structure [19]. More generally, our model predicts that static representational similarity will emerge in CA1 whenever there are correlations in EC inputs and/or targets, and will reflect frequencies of occurrence. These principles of learning mirror those demonstrated in many prior neural network models that use error-driven learning and overlapping, distributed representations (e.g. [36,47], see also [48]).

Our simulations are consistent with findings from recent human fMRI studies that used pattern similarity to assess changes in representations across voxels in the hippocampus. In a high-resolution study, we found that neural representations of items that were consistently paired in a continuous sequence became more similar in all subfields of the hippocampus [6], analogous to the settled response in our pair structure simulations. Similarly, in a study of temporal community structure, we found that hippocampal representations of items from the same community became more similar than those from different communities [5]. Although this study was not performed at high resolution, probabilistic segmentation suggested that the effect was most reliable in CA1, consistent with the stronger CA1 effects in our simulations. Finally, an associative inference study found the strongest pattern similarity for transitively associated pairs in CA1 [14]; direct pairs were presented only once in this study, corresponding to our simulations using one exposure (figure 4*b*).

Our model builds on two recent neural network models of the hippocampus. The first is Ketz *et al.* [13], who developed a technique for training the MSP online with error-driven learning. This allowed us to explore what kinds of representations emerge during training in the MSP, which was not possible in earlier versions of these models. Earlier models posited that MSP connections did not show appreciable (or any) learning on the timescale of a single experiment [49], but there is now evidence that the MSP does learn at this timescale [50–52]. The second is REMERGE [21], which demonstrated the utility of big-loop recurrence in transitive inference, paired associate inference, and acquired equivalence. Such recurrence plays three roles in our simulations: (i) it allows the structure learned in CA1 to spread to DG and CA3; (ii) it strengthens the similarity structure in CA1, due to the actual presence of the pairmate representation in EC_{in}; and (iii) when pairwise co-occurrence is not sufficient to uncover structure (i.e. community structure and associative inference), big-loop recurrence enhances transitive relationships. Like REMERGE, our model learns statistics while protecting pattern-separated representations used for episodic memory.

By contrast, however, our model outlines a role for static representational similarity in the hippocampus.

(a) Static representations versus recurrence

There is a debate in the hippocampal generalization literature as to whether transitive associations are supported by overlapping memory representations formed during encoding or by recurrent activity online at retrieval [20,21,32,42]. The core question is whether generalizations are stored in weights in the network or dynamically computed as needed. Our account is a hybrid between REMERGE, which depends entirely on recurrent dynamics, and models that depend entirely on overlap in pair representations [32,53]. In our model, recurrent dynamics play a dominant role very early in training, before the MSP has a chance to learn overlapping representations (this dependence on number of exposures has been proposed before, [20]). After sufficient exposure, however, both static representational similarity and recurrence contribute.

There is a further distinction to be made in possible learning mechanisms within the static representations account. The *integrative encoding* hypothesis [20] posits that viewing *BC* triggers retrieval of *AB* through pattern completion in the TSP. *AB* is then re-encoded in conjunction with the currently viewed *BC*, leading to an integrated static representation of *ABC* that can be used to support *AC* inference at test.

The learning occurring in our model's MSP is distinct from integrative encoding in several respects. First, whereas integrative encoding can occur, in theory, after one presentation of each pair, learning in the model's MSP is slower and requires multiple interleaved exposures (akin to cortex in CLS; [1]). Second, learning in our model's MSP, with sufficient exposure, does not depend on or require the TSP, as demonstrated by the TSP-lesion simulations. Third, learning in our model does not make use of overt reactivation of transitive associates (electronic supplementary material, figure S4). Relationships are uncovered implicitly as a result of the fact that patterns with shared features tend to adopt overlapping neural substrates in CA1. In other words, *AB* and *BC* become associated not because *BC* brought *AB* to mind, but simply because *AB* and *BC* share *B*. Speculatively, the rapid and overt reinstatement required for integrative encoding is perhaps more likely to occur in situations where pair encoding is made an explicit task requirement, as is the case in many associative inference but not statistical learning studies. When and how integrative encoding might occur, including whether it happens in the hippocampus alone or requires interactions with medial prefrontal cortex—an implicated area involved in processing event structure [19,54–56]—is an important area for future work.

How can these mechanisms—recurrence, static representations resulting from integrative encoding, and static representations resulting from implicit, interleaved learning—be experimentally disentangled? In the fMRI literature, increased hippocampal activity over training that correlates with generalization performance has been taken as evidence of integrative encoding in an acquired equivalence paradigm [57], and higher hippocampal activity for transitive pairs at test has been interpreted as flexible recombination of direct pairs at retrieval in associative inference [58]. However, REMERGE accounts for the acquired equivalence results using only recurrent dynamics [21], and our model can produce transitive associative inferences at test with static representational similarity (figure 4*d*,*e*).

Use of these mechanisms can be more directly revealed through lesion studies and by assessing fine-grained temporal dynamics in electrophysiological experiments with population recordings at different sites in the hippocampus and EC. If static representations of temporal regularities are acquired via interleaved learning (in the manner described by our model), then we can make several predictions about how these regularities will manifest after learning has taken place. Over the course of one trial, a single item should elicit a pattern of activity similar to that of its temporal associates first in CA1, then in deep layers of EC, then in superficial layers of EC, and finally in DG and CA3. The model therefore predicts that regularities will be eventually measureable in all subfields, but in a very particular order, with CA1 emerging first. Lesions to any portion of the TSP during or after learning should not affect behavioural evidence of statistical learning (after sufficient exposure), nor should it affect representations in CA1. Conversely, lesions to the MSP during or after learning should weaken representational and behavioural effects. Our model makes the counterintuitive prediction that an MSP lesion should leave in place weak sensitivity to pair structure early in learning, but that this ability should then deteriorate over time.

If generalization is based purely on recurrence (in the manner predicted by REMERGE, or by our model very early in training), then representations of associated items should arise during a trial only after activity has travelled from deep to superficial layers of EC. Given that REMERGE depends on strong pattern separation (though pattern separation and subfields were not directly implemented), it would predict deficits with a lesion to the TSP during or after learning.

If integrative encoding is used to learn temporal regularities, then—as in our model—representations of associated items should be present during the initial pass through the hippocampus, prior to activity spreading from deep to superficial layers of EC. Depending on the specific implementation of the mechanism (which has yet to be instantiated in a neural network model), the TSP may or may not be required for retrieval after learning is complete. However, in any implementation, the TSP would be required for encoding the integrated representations, and so lesions to the TSP should disrupt learning.

To summarize the two key predictions/differences: (i) if the learner is relying on recurrence or integrative encoding, TSP lesions should cause a learning deficit, whereas if the learner is relying on the mechanism described in our model (static representations acquired through interleaved learning), TSP lesions should not impair learning. It follows that infants, if using recurrence or integrative encoding, should have difficulty with transitive associations, whereas they should be able to learn adequately if using the mechanisms in our model. The insensitivity to TSP lesions in our model occurs after sufficient exposures to allow the MSP to learn; very early on in learning (e.g. after one exposure to each direct pair), our model depends on the TSP and recurrence, as in REMERGE. (ii) If regularities are expressed using static representations (via either the mechanism in our model or integrative encoding), representations of associated items should appear during the first pass through the hippocampal big loop, which will not be the case for recurrence. Different sets of these mechanisms may turn out to be

employed in different task contexts, at different periods over the course of learning, and/or by different individuals.

Note that the above predictions about recurrence and integrative encoding apply to paradigms involving transitive associations. It is unclear whether these two mechanisms could benefit pair structure learning, as they would memorize transitions between pairs in addition to the pairs themselves (as in our model's TSP), causing recurrence/integration to spread both within and across pairs. This limitation applies to other statistical learning paradigms (e.g. with triplet structure), and more generally to any paradigm with many repetitions that requires graded sensitivity to frequencies of co-occurrence. Simulations of these learning mechanisms applied directly to such paradigms will be useful.

(b) Anterior versus posterior hippocampal function

There is an anatomical gradient in the prominence of different hippocampal subfields, in which regions CA1–3 are relatively over-represented in anterior hippocampus and DG is over-represented in posterior hippocampus [59]. Considering DG as an important contributor to pattern separation in the TSP, this suggests there may be a relatively more dominant role of the MSP in the anterior hippocampus. Our model then makes the prediction that the anterior hippocampus should be more dominant in statistical learning. This is consistent with human fMRI findings on the anterior hippocampus: (i) it generally shows stronger effects in pair structure and community structure paradigms [5,6]; (ii) it plays a stronger role in integrating memories in the associative inference task [58,60]; (iii) it fully integrates across elements in an event narrative analogue of the associative inference task [61]; (iv) it is active during transitive inference [62]; and (v) its activity varies with generalization performance in acquired equivalence [57].

This generalization of experiences across time in anterior hippocampus may be related to the gradient of spatial representations found in rodent hippocampus, where ventral (anterior) place cells represent larger, more overlapping areas of space compared to dorsal (posterior) ones [61,63,64]. Analogous to our account, this varying spatial scale has been proposed to allow the hippocampus to carry out the complementary tasks of generalizing and avoiding interference in spatial memory [65]. Future extensions to our model could explore this relationship to spatial processing. With spatial information represented in EC [66], exposure to regularly occurring configurations, as has been studied in spatial statistical learning [67] and contextual cueing paradigms [68], would be expected to tap into the same learning mechanisms and pathways as the temporal cases. The MSP might integrate information more easily across larger areas of space, creating larger place fields, which would correspond to those found in anterior hippocampus.

(c) Temporal and sequence processing

There is substantial evidence from the rodent literature that the MSP and CA1 play a special role in temporal processing, including that: inhibition of the MSP leads to deficits in temporal association memory [69], CA1 supports temporal order memory [70–72], hippocampal place fields in CA1 expand to represent earlier positions on a track [73], and CA1 can generate reliable sequential spiking patterns on its own [74]. There is also evidence from human fMRI that CA1 plays a larger role than other hippocampal subfields in sequential processing [19,75].

These findings are consistent with our account of the MSP's role in extracting temporal regularities across experiences. They are surprising, however, given the large hippocampal modelling literature that has focused on the role of recurrent heteroassociative connections in CA3 in processing temporal sequences [30,76–80]. (See [81], though, for discussion of the consonant idea that less sparse representations in CA1 may allow it to represent sequences better than CA3.) One possibility is that greater reliance on the TSP versus MSP in sequence processing may depend on the type of sequence. If there is an obvious repeating sequence, the TSP may be best suited to learning and reproducing it, whereas a noisier sequence that requires integration across repetitions may require the MSP. It is also possible that big-loop recurrence or recurrence between EC and CA1 may play a similar computational role to recurrence within CA3. In our model (as in REMERGE), input *A* retrieves associated item *B*, which then travels through the big loop, such that the model first represents *A*, then additionally *B*. We tried training the model on a longer sequence (data not reported here), *ABCDEFGH*, and found that when given *A* at test, and with some synaptic depression, it would use the big loop to sequentially retrieve *B* (as *A* fades due to depression), then *C* (as *B* fades), then *D*, etc. This indicates that big-loop recurrence can be sufficient to produce robust sequential behaviour.

(d) Medial temporal lobe cortex and beyond

A related division of labour between episodic memory and statistical learning in the medial temporal lobe (MTL) has previously been proposed, with the hippocampus proper supporting episodic memory and EC learning incrementally and coming to represent co-occurring stimuli more similarly [82,83]. Empirical support for this hippocampus-EC dissociation comes from animal studies of incremental conditioning over days [84,85]. Although we did not model learning within EC, we expect that it will learn overlapping representations as in CA1, just on a slower timescale of days, not minutes. A slower learning rate in MTL cortex would still be consistent with it supporting rapid (even one-shot) familiarity for individual items [4,12], which can be instantiated as subtle changes to the representation of one item, as opposed to the more computationally difficult binding across multiple items. It would also be consistent with findings of rapid learning-related changes in representational similarity in MTL and other cortical areas [6]. This is because the hippocampus can reinstate an associated item in MTL simultaneously with processing of the current item in the same areas, causing apparent representational similarity without local learning (on this timescale).

One intriguing possibility is that there is a hierarchy of learning rates, perhaps related to a hierarchy of temporal window sizes for information accumulation [86], where the TSP is fastest, followed by the MSP, then MTL cortex, then cortical areas like the anterior temporal lobe that integrate across different types of sensory information [87], and then finally areas that support specific sensory functions. There is substantial evidence that the MSP/CA1 operates on a longer timescale than the TSP/CA3 [81,88–91], consistent with our model's relatively slower learning rate there. Further down the hierarchy, after days of learning to associate pairs of fractals, perirhinal cortex exhibits representational

similarity for paired fractals [92–94]. Such 'pair coding' neurons are found even further down in inferotemporal cortex (IT), but their response appears about 350 ms later than in perirhinal cortex [92], they are more prevalent in perirhinal cortex [93], and IT pair coding is abolished by perirhinal and entorhinal lesions [95]. For the rapid statistical learning considered here—on the order of minutes to hours—the learning rate of the MSP may be most suitable. The site of local learning may then expand outward with increasing exposure and opportunities for consolidation.

The idea that CA1 is intermediate in timescale between CA3 and MTL cortex may also relate to its role as a stable translator between CA3 and EC in episodic learning tasks. It is important for those mappings to be stable in order for CA1 to effectively communicate the details rapidly stored in DG and CA3 back to EC. However, this stability need only last while in a particular context (as opposed to over the lifespan, [49]).

These ideas also speak to a debate in the statistical learning literature regarding whether statistical learning is domain general or modality specific [96,97]. As argued above, statistical learning (in conventional tasks) occurs over a rapid timescale that is well-matched to the learning rate of the hippocampus and MSP specifically, but not to that of modality-specific sensory cortex. Nevertheless, processing dynamics, representational properties, and individual differences in these areas certainly shape the input to the hippocampus. Moreover, although the hippocampus is important for statistical learning within multiple modalities [4], it may use at least partially non-overlapping neural substrates to process inputs from each modality [98], such that the representations are not fully abstracted away from modality. Finally, as noted above for MTL cortex and IT, evidence of statistical learning could appear in modality-specific areas on a rapid timescale even without local learning, as a result of feedback from the hippocampus (e.g. input from visual cortex of image *A* allows the MSP to retrieve image *B* on EC_{out} and reinstate it back in visual cortex, likewise for statistical learning of a pair of sounds *AB* and auditory cortex). In sum, we propose that the hippocampus is a domain-general statistical learner constrained by and operating over specific modalities, and that it can influence corresponding modality-specific areas via reinstatement (which can themselves exhibit learning over a longer timescale).

(e) Conclusions and open questions

Where does this work leave us with respect to the original CLS theory, which posits that the hippocampus stores new information and slowly teaches it to cortex to prevent catastrophic interference [1]? We think that this fundamental principle remains unchallenged: it is still critical to use a separate memory store to protect long-term knowledge from novel, potentially interfering information, and we still think the hippocampus plays this role. The modification to the CLS framework that we propose is that both novel episodic and novel statistical information are quickly learned in the hippocampus—indeed, both types of information could interfere with cortical representations if learned directly in cortex. However, it is not the *entire* hippocampus that stores pattern-separated representations for distinct recent experiences, but rather this is the province of the TSP. DG and CA3 are the areas that truly instantiate the hippocampal learning properties as presented in CLS. The MSP, in contrast, specializes in extracting regularities across recent experiences. Indeed, CA1 is the most cortex-like area of the hippocampus, with more overlapping representations and a slower learning rate (though still sparser and faster than cortex)—precisely the properties that encourage neural networks to efficiently generalize across experiences.

One open question is whether the analogy to CLS consolidation might hold within this microcosm: on this short timescale, does the TSP train the MSP offline to help avoid catastrophic interference in the MSP? Another important question is how statistical representations in the MSP affect the process of consolidating information to cortex. CA1 neurons replay generalized trajectories in a maze, not just those directly experienced [99], raising the possibility that the representations in CA1 allow a more sophisticated and potentially more useful transfer of information to cortex (for related discussion and simulations, see [21]). Statistical learning in the hippocampus may also encourage replay of the most statistically reliable experiences, as opposed to indiscriminate replay of all experiences.

Future work could also explore whether the two pathways can be controlled depending on the *type* of past information that is currently relevant, or anticipated to be relevant in the future. For example, when trying to remember the specifics of a recent experience that overlaps with other recent experiences, perhaps activity in the MSP can be strategically suppressed. Future simulations could assess interactions between the pathways in the context of paradigms in which both episodic details and statistical information are salient or relevant to behaviour, such as experiments that demonstrate sampling of trial-unique information after learning statistical information across trials [100].

Finally, there have been recent findings that representations of statistically associated items can differentiate from one another (i.e. their neural patterns become *less* similar) in some situations [6,60,101], which may be due to competition induced among associated items. Future work will explore whether incorporating competition-dependent learning mechanisms [102–104] can allow the model to account for these findings.

Though there are many questions that remain to be addressed, we hope the model provides a useful step on the path towards understanding how the hippocampus supports memory for both specific episodes and the regularities that hold across them.

Data accessibility. The Emergent project file used for simulations is available at: http://arks.princeton.edu/ark:/88435/dsp010v838304r.

Authors' contributions. A.C.S. conceived the study, built and ran the models, analysed the data and wrote a draft of the manuscript. All authors contributed to the design of the study, discussed the results and edited the manuscript.

Competing interests. We have no competing interests.

Funding. This study is funded by NIH R01EY021755 (N.B.T.-B.); NIH R01MH069456 (K.A.N.); John Templeton Foundation (all authors).

Acknowledgements. The authors thank Samuel Ritter, who contributed to this project during a rotation in K.A.N.'s lab, and helpful conversations with Michael Arcaro, Roy Cox, Marc Howard, Nicholas Ketz, James McClelland, and Robert Stickgold.

Disclaimer. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

References

1. McClelland JL, McNaughton BL, O'Reilly RC. 1995 Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev*. **102**, 419–457. (doi:10.1037/0033-295X.102.3.419)
2. O'Reilly RC, Bhattacharyya R, Howard MD, Ketz N. 2014 Complementary learning systems. *Cogn. Sci*. **38**, 1229–1248. (doi:10.1111/j.1551-6709.2011.01214.x)
3. Schapiro AC, Turk-Browne NB. 2015 Statistical learning. In *Brain mapping: an encyclopedic reference* (ed. AW Toga), pp. 501–506. New York, NY: Academic Press.
4. Schapiro AC, Gregory E, Landau B, McCloskey M, Turk-Browne NB. 2014 The necessity of the medial temporal lobe for statistical learning. *J. Cogn. Neurosci*. **26**, 1736–1747. (doi:10.1162/jocn_a_00578)
5. Schapiro AC, Turk-Browne NB, Norman KA, Botvinick MM. 2016 Statistical learning of temporal community structure in the hippocampus. *Hippocampus* **26**, 3–8. (doi:10.1002/hipo.22523)
6. Schapiro AC, Kustner LV, Turk-Browne NB. 2012 Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr. Biol*. **22**, 1622–1627. (doi:10.1016/j.cub.2012.06.056)
7. Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM. 2010 Implicit perceptual anticipation triggered by statistical learning. *J. Neurosci*. **30**, 11 177–11 187. (doi:10.1523/JNEUROSCI.0858-10.2010)
8. Bornstein AM, Daw ND. 2012 Dissociating hippocampal and striatal contributions to sequential prediction learning. *Eur. J. Neurosci*. **35**, 1011–1023. (doi:10.1111/j.1460-9568.2011.07920.x)
9. Harrison LM, Duggins A, Friston KJ. 2006 Encoding uncertainty in the hippocampus. *Neural Netw*. **19**, 535–546. (doi:10.1016/j.neunet.2005.11.002)
10. Strange BA, Duggins A, Penny W, Dolan RJ, Friston KJ. 2005 Information theory, novelty and hippocampal responses: unpredicted or unpredictable? *Neural Netw*. **18**, 225–230. (doi:10.1016/j.neunet.2004.12.004)
11. Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK. 2009 Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J. Cogn. Neurosci*. **21**, 1934–1945. (doi:10.1162/jocn.2009.21131)
12. Norman KA, O'Reilly RC. 2003 Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychol. Rev*. **110**, 611–646. (doi:10.1037/0033-295X.110.4.611)
13. Ketz N, Morkonda SG, O'Reilly RC. 2013 Theta coordinated error-driven learning in the hippocampus. *PLoS Comput. Biol*. **9**, e1003067. (doi:10.1371/journal.pcbi.1003067)
14. Schlichting ML, Zeithamova D, Preston AR. 2014 CA1 subfield contributions to memory integration and inference. *Hippocampus* **24**, 1248–1260. (doi:10.1002/hipo.22310)
15. Fiser J, Aslin RN. 2002 Statistical learning of higher-order temporal structure from visual shape sequences. *J. Exp. Psychol. Learn. Mem. Cogn*. **28**, 458–467. (doi:10.1037/0278-7393.28.3.458)
16. Saffran JR, Aslin RN, Newport EL. 1996 Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928. (doi:10.1126/science.274.5294.1926)
17. Kirkham NZ, Slemmer JA, Johnson SP. 2002 Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* **83**, B35–B42. (doi:10.1016/S0010-0277(02)00004-5)
18. Gomez RL, Edgin JO. 2016 The extended trajectory of hippocampal development: implications for early memory development and disorder. *Dev. Cogn. Neurosci*. **18**, 57–69. (doi:10.1016/j.dcn.2015.08.009)
19. Schapiro AC, Rogers TT, Cordova NI, Turk-Browne NB, Botvinick MM. 2013 Neural representations of events arise from temporal community structure. *Nat. Neurosci*. **16**, 486–492. (doi:10.1038/nn.3331)
20. Zeithamova D, Schlichting ML, Preston AR. 2012 The hippocampus and inferential reasoning: building memories to navigate future decisions. *Front. Hum. Neurosci*. **6**, 70. (doi:10.3389/fnhum.2012.00070)
21. Kumaran D, McClelland JL. 2012 Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. *Psychol. Rev*. **119**, 573–616. (doi:10.1037/a0028681)
22. O'Reilly RC, Rudy JW. 2001 Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychol. Rev*. **108**, 311–345. (doi:10.1037/0033-295X.108.2.311)
23. Aisa B, Mingus B, O'Reilly R. 2008 The emergent neural modeling system. *Neural Netw*. **21**, 1146–1152. (doi:10.1016/j.neunet.2008.06.016)
24. O'Reilly RC, Munakata Y, Frank MJ, Hazy TE, Contributors. 2014 *Computational cognitive neuroscience*, 2nd edn., ch. 8. See https://grey.colorado.edu/CompCogNeuro/index.php/CCNBook/MainAugust.
25. O'Reilly RC, Munakata Y. 2000 *Computational explorations in cognitive neuroscience: understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.
26. O'Reilly RC. 1996 Biologically plausible error-driven learning using local activation differences: the generalized recirculation algorithm. *Neural Comput*. **8**, 895–938. (doi:10.1162/neco.1996.8.5.895)
27. Hasselmo ME, Bodelon C, Wyble BP. 2002 A proposed function for hippocampal theta rhythm: separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Comput*. **14**, 793–817. (doi:10.1162/089976602317318965)
28. Brankack J, Stewart M, Fox SE. 1993 Current source density analysis of the hippocampal theta rhythm: associated sustained potentials and candidate synaptic generators. *Brain Res*. **615**, 310–327. (doi:10.1016/0006-8993(93)90043-M)
29. Kumaran D, Maguire EA. 2006 An unexpected sequence of events: mismatch detection in the human hippocampus. *PLoS Biol*. **4**, e424. (doi:10.1371/journal.pbio.0040424)
30. Lisman JE. 1999 Relating hippocampal circuitry to function: recall of memory sequences by reciprocal dentate-CA3 interactions. *Neuron* **22**, 233–242. (doi:10.1016/S0896-6273(00)81085-5)
31. Suzuki WA, Miller EK, Desimone R. 1997 Object and place memory in the macaque entorhinal cortex. *J. Neurophysiol*. **78**, 1062–1081.
32. Howard MW, Fotedar MS, Datey AV, Hasselmo ME. 2005 The temporal context model in spatial navigation and relational learning: toward a common explanation of medial temporal lobe function across domains. *Psychol. Rev*. **112**, 75–116. (doi:10.1037/0033-295X.112.1.75)
33. Otto T, Eichenbaum H. 1992 Complementary roles of the orbital prefrontal cortex and the perirhinal-entorhinal cortices in an odor-guided delayed-nonmatching-to-sample task. *Behav. Neurosci*. **106**, 762–775. (doi:10.1037/0735-7044.106.5.762)
34. Murray EA, Mishkin M. 1998 Object recognition and location memory in monkeys with excitotoxic lesions of the amygdala and hippocampus. *J. Neurosci*. **18**, 6568–6582.
35. Mumby DG, Pinel JP. 1994 Rhinal cortex lesions and object recognition in rats. *Behav. Neurosci*. **108**, 11–18. (doi:10.1037/0735-7044.108.1.11)
36. Rogers TT, McClelland JL. 2004 *Semantic cognition: a parallel distributed processing approach*. Cambridge, MA: MIT Press.
37. Lisman JE, Redish AD. 2009 Prediction, sequences and the hippocampus. *Phil. Trans. R. Soc. B* **364**, 1193–1201. (doi:10.1098/rstb.2008.0316)
38. Buckner RL. 2010 The role of the hippocampus in prediction and imagination. *Annu. Rev. Psychol*. **61**, 27–48, C1–8. (doi:10.1146/annurev.psych.60.110707.163508)
39. Hindy NC, Ng FY, Turk-Browne NB. 2016 Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nat. Neurosci*. **19**, 665–667. (doi:10.1038/nn.4284)
40. Lavenex P, Banta Lavenex P. 2013 Building hippocampal circuits to learn and remember: insights into the development of human memory. *Behav. Brain Res*. **254**, 8–21. (doi:10.1016/j.bbr.2013.02.007)
41. Fortunato S. 2010 Community detection in graphs. *Phys. Rep. Rev. Sect. Phys. Lett*. **486**, 75–174. (doi:10.1016/j.physrep.2009.11.002)
42. Kumaran D. 2012 What representations and computations underpin the contribution of the hippocampus to generalization and inference? *Front. Hum. Neurosci*. **6**, 157. (doi:10.3389/fnhum.2012.00157)

43. Wood ER, Dudchenko PA, Robitsek RJ, Eichenbaum H. 2000 Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* **27**, 623–633. (doi:10.1016/S0896-6273(00)00071-4)
44. Singer AC, Karlsson MP, Nathe AR, Carr MF, Frank LM. 2010 Experience-dependent development of coordinated hippocampal spatial activity representing the similarity of related locations. *J. Neurosci*. **30**, 11 586–11 604. (doi:10.1523/JNEUROSCI.0926-10.2010)
45. Leutgeb JK, Leutgeb S, Moser MB, Moser EI. 2007 Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science* **315**, 961–966. (doi:10.1126/science.1135801)
46. Eichenbaum H. 1999 The hippocampus and mechanisms of declarative memory. *Behav. Brain Res*. **103**, 123–133. (doi:10.1016/S0166-4328(99)00044-3)
47. McClelland JL, Rumelhart DE. 1985 Distributed memory and the representation of general and specific information. *J. Exp. Psychol. Gen*. **114**, 159–197. (doi:10.1037/0096-3445.114.2.159)
48. Mareschal D, French RM. 2017 TRACX2: a connectionist autoencoder using graded chunks to model infant visual statistical learning. *Phil. Trans. R. Soc. B* **372**, 20160057. (doi:10.1098/rstb.2016.0057)
49. McClelland JL, Goddard NH. 1996 Considerations arising from a complementary learning systems perspective on hippocampus and neocortex. *Hippocampus* **6**, 654–665. (doi:10.1002/(SICI)1098-1063(1996)6:6<654::AID-HIPO8>3.0.CO;2-G)
50. Remondes M, Schuman EM. 2003 Molecular mechanisms contributing to long-lasting synaptic plasticity at the temporoammonic-CA1 synapse. *Learn. Mem*. **10**, 247–252. (doi:10.1101/lm.59103)
51. Remondes M, Schuman EM. 2002 Direct cortical input modulates plasticity and spiking in CA1 pyramidal neurons. *Nature* **416**, 736–740. (doi:10.1038/416736a)
52. Aksoy-Aksel A, Manahan-Vaughan D. 2013 The temporoammonic input to the hippocampal CA1 region displays distinctly different synaptic plasticity compared to the Schaffer collateral input *in vivo*: significance for synaptic information processing. *Front. Synaptic Neurosci*. **5**, 5. (doi:10.3389/fnsyn.2013.00005)
53. Howard MW, Jing B, Rao VA, Provyn JP, Datey AV. 2009 Bridging the gap: transitive associations between items presented in similar temporal contexts. *J. Exp. Psychol. Learn. Mem. Cogn*. **35**, 391–407. (doi:10.1037/a0015002)
54. Krueger F, Moll J, Zahn R, Heinecke A, Grafman J. 2007 Event frequency modulates the processing of daily life activities in human medial prefrontal cortex. *Cereb. Cortex* **17**, 2346–2353. (doi:10.1093/cercor/bhl143)
55. Zeithamova D, Dominick AL, Preston AR. 2012 Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron* **75**, 168–179. (doi:10.1016/j.neuron.2012.05.010)
56. Schlichting ML, Preston AR. 2015 Memory integration: neural mechanisms and implications for behavior. *Curr. Opin. Behav. Sci*. **1**, 1–8. (doi:10.1016/j.cobeha.2014.07.005)
57. Shohamy D, Wagner AD. 2008 Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron* **60**, 378–389. (doi:10.1016/j.neuron.2008.09.023)
58. Preston AR, Shrager Y, Dudukovic NM, Gabrieli JD. 2004 Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus* **14**, 148–152. (doi:10.1002/hipo.20009)
59. Malykhin NV, Lebel RM, Coupland NJ, Wilman AH, Carter R. 2010 *In vivo* quantification of hippocampal subfields using 4.7 T fast spin echo imaging. *Neuroimage* **49**, 1224–1230. (doi:10.1016/j.neuroimage.2009.09.042)
60. Schlichting ML, Mumford JA, Preston AR. 2015 Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nat. Commun*. **6**, 8151. (doi:10.1038/ncomms9151)
61. Collin SH, Milivojevic B, Doeller CF. 2015 Memory hierarchies map onto the hippocampal long axis in humans. *Nat. Neurosci*. **18**, 1562–1564. (doi:10.1038/nn.4138)
62. Heckers S, Zalesak M, Weiss AP, Ditman T, Titone D. 2004 Hippocampal activation during transitive inference in humans. *Hippocampus* **14**, 153–162. (doi:10.1002/hipo.10189)
63. Poppenk J, Evensmoen HR, Moscovitch M, Nadel L. 2013 Long-axis specialization of the human hippocampus. *Trends Cogn. Sci*. **17**, 230–240. (doi:10.1016/j.tics.2013.03.005)
64. Howard MW, MacDonald CJ, Tiganj Z, Shankar KH, Du Q, Hasselmo ME, Eichenbaum H. 2014 A unified mathematical framework for coding time, space, and sequences in the hippocampal region. *J. Neurosci*. **34**, 4692–4707. (doi:10.1523/JNEUROSCI.5808-12.2014)
65. Keinath AT, Wang ME, Wann EG, Yuan RK, Dudman JT, Muzzio IA. 2014 Precise spatial coding is preserved along the longitudinal hippocampal axis. *Hippocampus* **24**, 1533–1548. (doi:10.1002/hipo.22333)
66. Moser EI, Kropff E, Moser MB. 2008 Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci*. **31**, 69–89. (doi:10.1146/annurev.neuro.31.061307.090723)
67. Fiser J, Aslin RN. 2001 Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol. Sci*. **12**, 499–504. (doi:10.1111/1467-9280.00392)
68. Chun MM, Phelps EA. 1999 Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nat. Neurosci*. **2**, 844–847. (doi:10.1038/12222)
69. Suh J, Rivest AJ, Nakashiba T, Tominaga T, Tonegawa S. 2011 Entorhinal cortex layer III input to the hippocampus is crucial for temporal association memory. *Science* **334**, 1415–1420. (doi:10.1126/science.1210125)
70. Hoge J, Kesner RP. 2007 Role of CA3 and CA1 subregions of the dorsal hippocampus on temporal processing of objects. *Neurobiol. Learn. Mem*. **88**, 225–231. (doi:10.1016/j.nlm.2007.04.013)
71. Manns JR, Howard MW, Eichenbaum H. 2007 Gradual changes in hippocampal activity support remembering the order of events. *Neuron* **56**, 530–540. (doi:10.1016/j.neuron.2007.08.017)
72. Allen TA, Salz DM, McKenzie S, Fortin NJ. 2016 Nonspatial sequence coding in CA1 neurons. *J. Neurosci*. **36**, 1547–1563. (doi:10.1523/JNEUROSCI.2874-15.2016)
73. Mehta MR, Barnes CA, McNaughton BL. 1997 Experience-dependent, asymmetric expansion of hippocampal place fields. *Proc. Natl Acad. Sci. USA* **94**, 8918–8921. (doi:10.1073/pnas.94.16.8918)
74. Stark E, Roux L, Eichler R, Buzsaki G. 2015 Local generation of multineuronal spike sequences in the hippocampal CA1 region. *Proc. Natl Acad. Sci. USA* **112**, 10 521–10 526. (doi:10.1073/pnas.1508785112)
75. Wang F, Diana RA. 2016 Temporal context processing within hippocampal subfields. *Neuroimage* **134**, 261–269. (doi:10.1016/j.neuroimage.2016.03.048)
76. Levy WB. 1996 A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus* **6**, 579–590. (doi:10.1002/(SICI)1098-1063(1996)6:6<579::AID-HIPO3>3.0.CO;2-C)
77. Wallenstein GV, Hasselmo ME. 1997 GABAergic modulation of hippocampal population activity: sequence learning, place field development, and the phase precession effect. *J. Neurophysiol*. **78**, 393–408.
78. Jensen O, Lisman JE. 1996 Hippocampal CA3 region predicts memory sequences: accounting for the phase precession of place cells. *Learn. Mem*. **3**, 279–287. (doi:10.1101/lm.3.2-3.279)
79. Tsodyks MV, Skaggs WE, Sejnowski TJ, McNaughton BL. 1996 Population dynamics and theta rhythm phase precession of hippocampal place cell firing: a spiking neuron model. *Hippocampus* **6**, 271–280. (doi:10.1002/(SICI)1098-1063(1996)6:3<271::AID-HIPO5>3.3.CO;2-Q)
80. Rolls ET. 2013 A quantitative theory of the functions of the hippocampal CA3 network in memory. *Front. Cell Neurosci*. **7**, 98. (doi:10.3389/fncel.2013.00098)
81. Rolls ET, Kesner RP. 2006 A computational theory of hippocampal function, and empirical tests of the theory. *Prog. Neurobiol*. **79**, 1–48. (doi:10.1016/j.pneurobio.2006.04.005)
82. Myers CE, Gluck MA, Granger R. 1995 Dissociation of hippocampal and entorhinal function in associative learning: a computational approach. *Psychobiology* **23**, 116–138.
83. Gluck MA, Meeter M, Myers CE. 2003 Computational models of the hippocampal region: linking incremental learning and episodic memory. *Trends Cogn. Sci*. **7**, 269–276. (doi:10.1016/S1364-6613(03)00105-0)

84. Allen TA, Chelius L, Gluck MA. 2002 Selective entorhinal and nonselective cortical-hippocampal region lesions, but not selective hippocampal lesions, disrupt learned irrelevance in rabbit eyeblink conditioning. *Cogn. Affect. Behav. Neurosci*. **2**, 214–226. (doi:10.3758/CABN.2.3.214)
85. Shohamy D, Allen MT, Gluck MA. 2000 Dissociating entorhinal and hippocampal involvement in latent inhibition. *Behav. Neurosci*. **114**, 867–874. (doi:10.1037/0735-7044.114.5.867)
86. Hasson U, Chen J, Honey CJ. 2015 Hierarchical process memory: memory as an integral component of information processing. *Trends Cogn. Sci*. **19**, 304–313. (doi:10.1016/j.tics.2015.04.006)
87. Patterson K, Nestor PJ, Rogers TT. 2007 Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci*. **8**, 976–987. (doi:10.1038/nrn2277)
88. Gall CM, Hess US, Lynch G. 1998 Mapping brain networks engaged by, and changed by, learning. *Neurobiol. Learn. Mem*. **70**, 14–36. (doi:10.1006/nlme.1998.3835)
89. Lee I, Rao G, Knierim JJ. 2004 A double dissociation between hippocampal subfields: differential time course of CA3 and CA1 place cells for processing changed environments. *Neuron* **42**, 803–815. (doi:10.1016/j.neuron.2004.05.010)
90. Lee I, Kesner RP. 2002 Differential contribution of NMDA receptors in hippocampal subregions to spatial working memory. *Nat. Neurosci*. **5**, 162–168. (doi:10.1038/nn790)
91. Nakashiba T, Young JZ, McHugh TJ, Buhl DL, Tonegawa S. 2008 Transgenic inhibition of synaptic transmission reveals role of CA3 output in hippocampal learning. *Science* **319**, 1260–1264. (doi:10.1126/science.1151120)
92. Naya Y, Yoshida M, Miyashita Y. 2001 Backward spreading of memory-retrieval signal in the primate temporal cortex. *Science* **291**, 661–664. (doi:10.1126/science.291.5504.661)
93. Naya Y, Yoshida M, Miyashita Y. 2003 Forward processing of long-term associative memory in monkey inferotemporal cortex. *J. Neurosci*. **23**, 2861–2871.
94. Erickson CA, Desimone R. 1999 Responses of macaque perirhinal neurons during and after visual stimulus association learning. *J. Neurosci*. **19**, 10 404–10 416.
95. Miyashita Y, Kameyama M, Hasegawa I, Fukushima T. 1998 Consolidation of visual associative long-term memory in the temporal cortex of primates. *Neurobiol. Learn. Mem*. **70**, 197–211. (doi:10.1006/nlme.1998.3848)
96. Frost R, Armstrong BC, Siegelman N, Christiansen MH. 2015 Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn. Sci*. **19**, 117–125. (doi:10.1016/j.tics.2014.12.010)
97. Siegelman N, Bogaerts L, Christiansen MH, Frost R. 2017 Towards a theory of individual differences in statistical learning. *Phil. Trans. R. Soc. B* **372**, 20160059. (doi:10.1098/rstb.2016.0059)
98. Hunsaker MR, Fieldsted PM, Rosenberg JS, Kesner RP. 2008 Dissociating the roles of dorsal and ventral CA1 for the temporal processing of spatial locations, visual objects, and odors. *Behav. Neurosci*. **122**, 643–650. (doi:10.1037/0735-7044.122.3.643)
99. Gupta AS, van der Meer MA, Touretzky DS, Redish AD. 2010 Hippocampal replay is not a simple function of experience. *Neuron* **65**, 695–705. (doi:10.1016/j.neuron.2010.01.034)
100. Bornstein AM, Khaw MW, Shohamy D, Daw ND. 2015 What's past is present: reminders of past choices bias decisions for reward in humans. *bioRxiv* (doi:10.1101/033910)
101. Favila SE, Chanales AJH, Kuhl AK. 2016 Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nat. Commun*. **7**, 11066. (doi:10.1038/ncomms11066)
102. Norman KA, Newman EL, Detre G. 2007 A neural network model of retrieval-induced forgetting. *Psychol. Rev*. **114**, 887–953. (doi:10.1037/0033-295X.114.4.887)
103. Norman KA, Newman E, Detre G, Polyn S. 2006 How inhibitory oscillations can train neural networks and punish competitors. *Neural Comput*. **18**, 1577–1610. (doi:10.1162/neco.2006.18.7.1577)
104. Hulbert JC, Norman KA. 2015 Neural differentiation tracks improved recall of competing memories following interleaved study and retrieval practice. *Cereb. Cortex* **25**, 3994–4008. (doi:10.1093/cercor/bhu284)