Received: 3 May 2019 | Revised: 28 November 2019 | Accepted: 4 December 2019

DOI: 10.1002/hipo.23186

RESEARCH ARTICLE

WILEY

Improving sensory representations using episodic memory

Richard Görler[1,2] | Laurenz Wiskott[1] | Sen Cheng[1]

[1]Institute for Neural Computation, Ruhr University Bochum, Bochum, Germany

[2]International Graduate School of Neuroscience, Ruhr University Bochum, Bochum, Germany

Correspondence
Sen Cheng, Institute for Neural Computation, Ruhr University Bochum, Bochum 44801, Germany.
Email: sen.cheng@rub.de;

Funding information
Bundesministerium für Bildung und Forschung, Grant/Award Number: 01GQ1506; Deutsche Forschungsgemeinschaft, Grant/Award Numbers: 419039588, 122679504

Abstract

The medial temporal lobe (MTL) is well known to be essential for declarative memory. However, a growing body of research suggests that MTL structures might be involved in perceptual processes as well. Our previous modeling work suggests that sensory representations in cortex influence the accuracy of episodic memory retrieved from the MTL. We adopt that model here to show that, conversely, episodic memory can also influence the quality of sensory representations. We model the effect of episodic memory as (a) repeatedly replaying episodes from memory and (b) recombining episode fragments to form novel sequences that are more informative for learning sensory representations than the original episodes. We demonstrate that the performance in visual discrimination tasks is superior when episodic memory is present and that this difference is due to episodic memory driving the learning of a more optimized sensory representation. We conclude that the MTL can, even if it has only a purely mnemonic function, influence perceptual discrimination indirectly.

KEYWORDS

discrimination (psychology), learning, memory consolidation, models, theoretical, temporal lobe

1 | INTRODUCTION

We use episodic memory to remember events that we have experienced ourselves (Tulving, 1972). However, while we may remember the basic events and their sequential relation in an episode, we cannot recall the detailed sensory information that we experienced during the episode. Indeed, experimental studies have found that episodic memory in humans preserves mostly the gist of the experienced episode and few of the details (Koutstaal & Schacter, 1997; Sachs, 1967). We have suggested previously that this property of the episodic memory system results from the fact that episodes are stored in terms of a higher order representation of sensory input and not the sensory input itself (Cheng, Werning, & Suddendorf, 2016; Fang, Demic, & Cheng, 2018; Fang, Rüther, Bellebaum, Wiskott, & Cheng, 2018), which is compatible with the indexing theory by Teyler and DiScenna (1986). Moreover, this is reminiscent of Tulving's SPI model, which posits that sensory information has to pass through the semantic system before being stored by the episodic system (Tulving, 1995). SPI stands for serial encoding, parallel storage and independent retrieval, which describes the phase-dependent information flow among the perceptual, semantic, and episodic components.

This higher order representation is generated during the processing of sensory information, as the information is high-dimensional and has to be represented by patterns of neural activity in cortex. Due to biophysical constraints (Ganguli & Sompolinsky, 2012), for example, metabolic costs (Lennie, 2003) or the space required for neuronal connections, the dimensionality of the cortical representation is reduced along the stream of sensory processing (Beyeler, Rounds, Carlson, Dutt, & Krichmar, 2019). At every stage of processing the sensory representation has to contain meaningful features that are informative of the content in the input. The nature of these features can be learned from statistical regularities in the input.

For example, consider the sensory representation of the visual system, which has been shown to process information hierarchically (Felleman & Van Essen, 1991). In primary visual cortex, input is represented by simple and complex cells (Hubel & Wiesel, 1959), hence represented by location, orientation and spatial frequency of elements in view. On the other end of the abstraction spectrum, in inferior temporal areas, shape, and identity of entire objects are represented (Gross, 1992).

The higher-level visual areas directly project to medial temporal lobe (MTL) structures, which are well known to be crucial for episodic memory (Scoville & Milner, 1957; Squire & Zola, 1998). We have previously suggested that episodic memories are best represented by sequences of neural activity patterns (Cheng, 2013; Cheng & Werning, 2016; Cheng et al., 2016). This representational format seems well supported by the hippocampus, which has been shown to be important for storing and retrieving sequences (Agster, Fortin, & Eichenbaum, 2002; Fortin, Agster, & Eichenbaum, 2002). Specifically, previous modeling work has suggested that the recurrent network in the hippocampal region CA3 encodes sequences during the experience and replays them during retrieval (Bayati et al., 2018; Buhry, Azizi, & Cheng, 2011; Cheng, 2013; Levy, 1996; Lisman, 1999). In summary, we have suggested that a perceptual-semantic representational network in the neocortex provides higher order representations of the sensory information, while the episodic memory trace only stores the gist of scenes and their temporal evolution.

According to the traditional view, the MTL exclusively subserves mnemonic processes (Squire, Stark, & Clark, 2004; Squire & Zola-Morgan, 1991). However, a growing body of results from the last two decades suggests that the MTL may also play a critical role in high-level perception (perceptual-mnemonic hypothesis; Buckley, Booth, Rolls, & Gaffan, 2001; Bussey & Saksida, 2007). Most of these studies apply variants of two basic types of perceptual tasks in lesion and fMRI experiments to test that hypothesis (Graham, Barense, & Lee, 2010).

1. Discrimination tasks, in which participants have to compare images and judge similarity. Lee et al. applied a morphing technique to generate image pairs with five different similarity levels containing faces, objects, natural scenes, or art (A. C. H. Lee, Fischer, et al., 2005), but see (Shrager, Gold, Hopkins, & Squire, 2006). Participants had to decide which of the two images is more similar to a reference image. Patients with broad MTL damage (including hippocampus and perirhinal cortex) were strongly impaired in scene discrimination and less so in face and object discrimination. Patients with specific hippocampal damage were only impaired in scene discrimination and only slightly.
2. Oddity judgment tasks, in which the participant has to pick the odd-one-out of a number of shown objects. These objects can be simple geometrical shapes, faces, familiar or novel objects, artificial scenes, often shown from different angles. Buckley et al. (2001) conducted such an oddity judgment study with monkeys and found that subjects with perirhinal cortex lesions were impaired. While Stark and Squire (2000) were not able to replicate this result in humans, Lee, Buckley et al. (2005) found similar impairments in patients with MTL damage, especially when the stimuli were shown from differing viewing angles.

A different task worth mentioning was used in a study by A. C. H. Lee and Rudebeck (2010), in which participants had to judge whether or not a line drawing of a novel object is geometrically possible. The results show that, firstly, a patient with broad MTL lesions performed poorer on the task than controls. Secondly, fixation patterns of the MTL patient, when responding incorrectly, differed from those of controls and hence the authors concluded that a deficit of visual processing and not a memory deficit is responsible.

Overall, the studies on MTL function in visual perception have been interpreted to suggest that the perirhinal cortex is involved in the visual perception of complex objects and faces by processing complex conjunctions of features, and that the hippocampus is involved in the visual processing of scenes, although there are alternative theories and contradicting evidence.

Based on results of a computational study and a preliminary experiment, it has recently been suggested that episodic memory retrieval is facilitated by an appropriate sensory representation (Fang, Rüther, et al., 2018). We adopt this model and propose that, conversely, episodic memory also leads to more optimized sensory representations. We do not aim to provide a detailed model of the MTL memory system, but focus on how episodic memory can serve to indirectly improve perception. We hypothesize that the sensory representations are initially learned through sensory experience but can be improved further by replaying experiences from memory, perhaps during a process of systems consolidation (Cheng, 2017). Using these sensory representations, we model a simple visual discrimination task and show that, after training with episodic memory, performance is better than without episodic memory. We discuss the modeling results with respect to the studies mentioned above, and explain how the model can account for experimental results.

2 | METHODS

The model consists of three components: Sensory input, the representational system, and the episodic memory system. In the following, we first give a brief overview of the model and then provide more details about the individual components below. For the sensory input, we use a stream of images. We use slow feature analysis (SFA; Wiskott & Sejnowski, 2002) to train the representational system to extract more abstract representations of the input images. SFA is an unsupervised learning algorithm that extracts slowly varying features (e.g., identity and position of an object in the input) from quickly varying data (e.g., pixel values). SFA training is based on changes of the input in time, but the learned feature representation is extracted from a single input image by an instantaneous function. This is consistent with the assumption in our modeling framework, that semantic memory is represented as near-instantaneous patterns of neural activity, as opposed to episodic memory, which is represented as sequences of activity patterns that are stored in hippocampus (Cheng, 2013; Cheng & Werning, 2016).

A sequential episodic memory system can be used to perform at least two manipulations on the original data: (a) Episodes can be replayed verbatim from memory multiple times. While the resulting data do not contain more information than the original data, learning

systems can profit from a simple repetition of the training data. Indeed, neural activity in the rat hippocampus has been shown to replay previously experienced sequences during sleep (A. K. Lee & Wilson, 2002; Louie & Wilson, 2001) or even during periods of rest in the awake state (Diba & Buzsáki, 2007; Foster & Wilson, 2006). Furthermore, there is evidence for coordinated replay in hippocampus and visual cortex (Ji & Wilson, 2007). (b) Episode fragments are recombined to form novel sequences that have smoother transitions than the original episodes. It is well conceivable that the retrieval of memories leads to the retrieval of other memories that are related. Indeed, offline hippocampal sequences corresponding to never-experienced paths have been observed in experiments (Gupta, van der Meer, Touretzky, & Redish, 2010). These novel paths were stitched together from fragments that individually were experienced before. We hypothesize that the novel, recombined sequences are more informative for learning sensory representations than the original episodes.

While in a biological system repeated replay and the generation of novel sequences certainly are intermingled, we model both processes separately in order to disambiguate the effects. In the replay condition, the total number of training patterns is increased by repetition and the original episodes are not altered. In the novel sequences condition, the total number of patterns is held constant, but their sequential ordering is different from the sequences in the stored episodes.

We do not store sensory input directly in episodic memory but a lower-dimensional representation of it (SFA_{lo}, Figure 1). To study the potential effect of episodic memory on tuning sensory representations, we use a second layer of sensory representation (SFA_{hi}, Figure 1) that processes the output of SFA_{lo}. The structure of the representational system in our model can be viewed as a simplified model of a hierarchical visual system. Before the actual experiment takes place, SFA_{lo} is pretrained (double dashed line in Figure 1) and then fixed for the remainder of the study. Then SFA_{hi} is trained on a set of training data. These training episodes are first fed through SFA_{lo}, with subsequent whitening, and the resulting data are used to train SFA_{hi} (dashed lines in Figure 1).

There are two different instances of SFA_{hi}, which we will compare. One instance is trained directly on the output of SFA_{lo} ($SFA_{hi}[S]$—"simple") and the other on sequences that were stored and retrieved from episodic memory ($SFA_{hi}[E]$—"episodic") after passing through SFA_{lo}. This notation is used for both memory models, verbatim replay and generation of novel sequences. After training SFA_{hi}, a set of test data is used to assess the quality of the sensory representation that the two SFA_{hi} instances have extracted (solid lines in Figure 1). The assessment criteria are described further below (Section 2.4). A different set of test data is used to simulate a visual discrimination task.

Whitening, which is performed on the output of SFA_{lo}, is a linear transformation that normalizes the data to have zero mean and variance one in all directions. The output of SFA_{lo} on the pretraining data are already whitened (Equations 2 and 3). However, because the training data are slightly different from the pretraining data, the output of

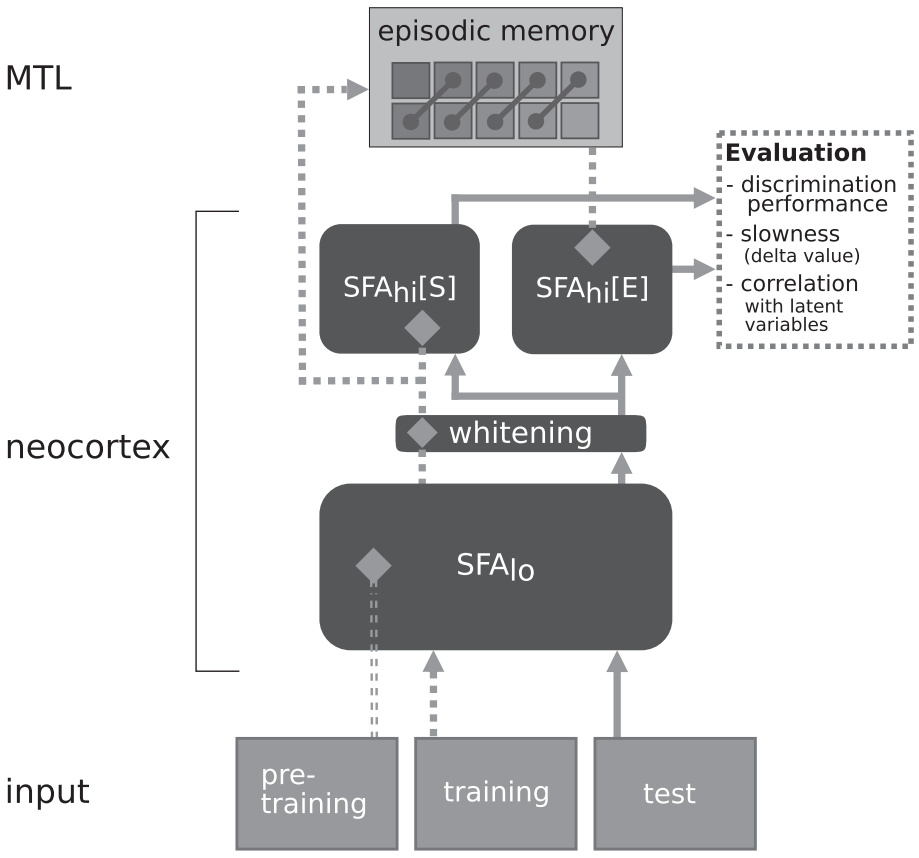

FIGURE 1 Structure of the representational system. The diagram depicts the information flow in the model, illustrating how each of the three data sets (bottom) are used in the three different stages of the simulation. A rhombus at the end of a line denotes that the data are used for training the particular module. An arrowhead indicates that the data are fed through the module. *Pretraining*: Before the actual experiments, SFA_{lo} is pretrained (double dashed line). *Training*: Training data are fed through SFA_{lo}, which extracts low-level visual features that are used to train SFA_{hi} (dashed lines). In our study, we contrast two SFA_{hi} instances. In the simple scenario $SFA_{hi}[S]$ is trained on the output of SFA_{lo} directly, while in the episodic scenario data are stored in episodic memory first and then retrieved to train $SFA_{hi}[E]$. *Test*: Finally, the quality of the features the SFA_{hi} instances extract is evaluated by feeding a set of testing data first through SFA_{lo} and then through SFA_{hi} (solid lines). For the role of the whitening see the main text

SFA_{lo} can have a mean and variance different from zero and one, respectively, when fed with training data. In our experiments we found that training of incremental SFA worked more consistently on whitened data, so we included the additional whitening step. The whitening matrix is trained on the output of SFA_{lo} on the training data and the same whitening matrix is used during testing.

2.1 | Sensory input

As input, we use streams of grayscale images with 30×30 pixels containing a single object, which is either the letter "T" or the letter "L". The black objects with smoothed edges are moving and rotating on a white background according to a random walk. The square images are represented in *x*- and *y*-coordinates ranging from −1 to 1. Changes in position and angle are drawn from a Gaussian distribution with zero mean and a standard deviation of 0.25. Independent Gaussian noise with zero mean and *SD* 0.1 is added to each pixel in each frame. Pixel values range from 0 to 1.

Each data set consists of several episodes of same length that are strung together and presented as one long stream (Figure 2). For each episode, the starting position and angle are randomly initialized. After each episode, the object identity changes, that is, the two letters are presented alternately (Figure 2). In the following, we refer to object identity and *x*,*y*-coordinates as latent variables. While the pixel values themselves can be directly observed, latent variables can only be inferred from pixel values.

Pre-training and test data consist of 100 episodes of length 50. The length and number of episodes in the training data varies between the experiments.

2.2 | Sensory representation

A well supported hypothesis about the function of the visual system is that it generates representations of the visual inputs that are invariant to many transformations, for example, changes of object orientation and position, view angle, lighting, and so on (Logothetis & Sheinberg, 1996; Rolls, 2000). SFA proposes that invariant representations are most likely those that are varying slowly in time. Therefore, the objective of SFA is to extract features from the input that vary as slowly as possible. Indeed, SFA applied to sequences of moving objects naturally learns a compact representation of object identity and pose (Franzius, Wilbert, & Wiskott, 2011), akin to that found in inferior temporal cortex. Experimental evidence supports the hypothesis that slowness may play an important role in forming neural representations (Li & DiCarlo, 2010). While the features being extracted by SFA on such a high level are usually easy to analyze and interpret, namely object identity and pose, the functions that extract the features are much less accessible. However, on the lowest level of the visual hierarchy, the functions can be analyzed and have been found to correspond to complex cell receptive fields (Berkes & Wiskott, 2005). Here we use a linear version of this model to learn plausible invariant representations of the visual input. Note that we model the representations formed in the neocortex with SFA, not the hippocampal mechanisms, although studies have demonstrated slowly changing neural signals (Cai et al., 2016; Tsao et al., 2018) as well as time cells (MacDonald, Lepage, Eden, & Eichenbaum, 2011; Pastalkova, Itskov, Amarasingham, & Buzsáki, 2008) in the hippocampus. As opposed to the representations generated by SFA, these are attributable to a putative explicit representation of time in the hippocampus.

SFA finds instantaneous scalar functions that generate slowly varying output from quickly varying input. Given a multidimensional input $x(t)$ and a function space F, SFA finds a set of functions $\{g^{(1)}(x), g^{(2)}(x), \ldots, g^{(i)}(x), \ldots\}$ with $g^{(i)}(x) \in F$, such that the output signals $y^{(i)}(t) := g^{(i)}(x(t))$, $\forall i$, minimize

$$\Delta\left(y^{(i)}\right) := \left\langle \left(\dot{y}^{(i)}\right)^2 \right\rangle_t \text{ (delta value)} \tag{1}$$

under the following constraints:

$$\left\langle y^{(i)} \right\rangle_t = 0 \text{ (zero mean)}, \tag{2}$$

$$\left\langle \left(y^{(i)}\right)^2 \right\rangle_t = 1 \text{ (unit variance)}, \tag{3}$$

$$\left\langle y^{(i)} y^{(j)} \right\rangle_t = 0 \; \forall j < i \text{ (decorrelation and order)}. \tag{4}$$

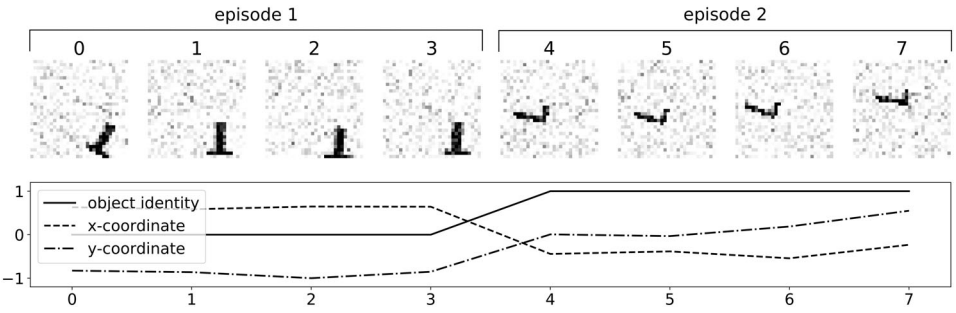

FIGURE 2 Example input. Shown are two episodes containing four images each (top), and the corresponding relevant latent variables (object identity and *x*,*y*-coordinates, bottom). Episodes are strung together to form one data set. The object is switched and its position is randomized at the start of each episode, hence the latent variables exhibit a jump at the transition between episodes

The delta value defined by Equation (1) is a measure of the slowness of the signal $y^{(i)}(t)$. It will also be used as one of the criteria for comparing the quality of different sensory representations. Equations (2) and (3) ensure that SFA does not generate the trivial solution of a constant function (for which $\Delta = 0$). The constraint in Equation (4) ensures that SFA does not yield the same feature twice and that the features are ordered according to the degree of their slowness, that is, $y^{(1)}(t)$ has the smallest delta value and $\Delta(y^{(i)}) < \Delta(y^{(j)})$ for $i < j$.

The standard batch implementation of SFA is available from the Python library Modular Toolkit for Data Processing (MDP; Zito, Wilbert, Wiskott, & Berkes, 2008). There is also an incremental version of the algorithm (Kompella, Luciw, & Schmidhuber, 2012), which has been shown to asymptotically reach the same performance as batch SFA with sufficient training and has the advantage of being a more plausible model for a biological learning system. In an incremental learning system, samples from the training data are presented one by one and the model parameters are updated every time a sample is presented. No memory of the previous samples is available, only the information stored in the model parameters of the learning system. By contrast, all samples are available at the same time in a batch learning algorithm and training is done once on the entire training data set. Since we want to study the effects of memory on the learning of representations, we have to dissociate memory from the learning process. Hence, for the high-level sensory representation (SFA_{hi}) we use the incremental algorithm which does not require holding the entire training data set in memory. The learning rate is set to 0.005, which yielded the best asymptotic performance in our scenarios. However, because the low-level representation (SFA_{lo}) serves only preprocessing purposes and remains unchanged during the experiments, SFA_{lo} is implemented with the batch algorithm to be sure to reach optimal performance.

For simplicity, we operate in a linear function space ("linear SFA"), which is sufficient to reliably extract the features of interest (position and identity of the object). We use several identical SFA_{lo} nodes with overlapping receptive fields as a simple model of receptive fields in visual cortex, as it is common in work on SFA. As illustrated in Figure 3, the receptive field of each node of SFA_{lo} spans an 18×18 pixel area of the input image and has an overlap of 6 pixels with the receptive fields of the neighboring nodes. Thus, 3×3 nodes jointly cover the image space. Each node of SFA_{lo} generates 32 features, and all 9×32 SFA_{lo} features are strung together in a single vector. These features are further processed by SFA_{hi}, which is a single node of linear incremental SFA that extracts 16 features.

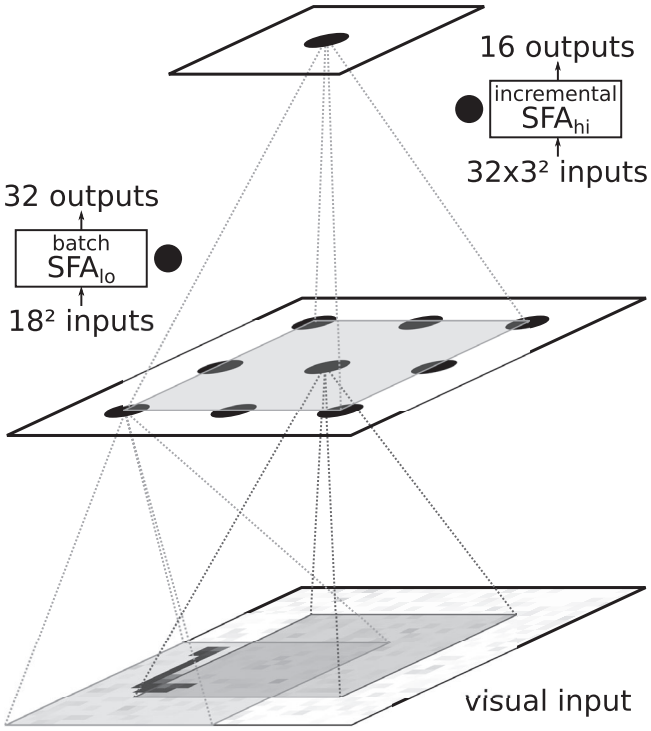

FIGURE 3 Structure of the slow feature analysis (SFA) network. The black dots represent SFA nodes. The gray patches represent the receptive fields that partially overlap in the SFA_{lo} network

2.3 | Episodic memory

The training data are stored in episodic memory after passing through SFA_{lo}. The focus of our model is the function of a sequential episodic memory in the formation of an optimal sensory representation and not the functioning of the memory system itself. Therefore, our episodic memory model is highly simplified. Its sequential nature is reminiscent of theoretical considerations proposing that the hippocampus holds a record of past stimuli and episodic memory recall results in a "jump back in time" (Howard & Eichenbaum, 2013). Experimental data indeed show that the population vector in the hippocampus changes gradually over time and on successful memory retrieval, the population activity at the time of encoding is reinstated (Folkerts, Rutishauser, & Howard, 2018). While the population vector in the hippocampus might encode past, present and even future to some extent at the same time, our model is simplified to focus on the sequential nature of episodic memory without an explicit representation of time or context. However, recalling a past memory reinstates the activity (= the retrieved pattern) that was present at the time of experience in the model as well. Furthermore, since the stored patterns are SFA features, and these features change slowly in time, the patterns in the hippocampus change gradually at the time of experience.

When modeling storage and retrieval of episodic memory, we distinguish two different modes as mentioned above.

2.3.1 | Repeated verbatim replay of the episodes

In the simple scenario SFA_{hi}[S] is trained on the training data consisting of 100 episodes of length 50. Episodic training is modeled by repeated training of SFA_{hi}[E] on the same data set for up to 40 iterations. In each repetition, every episode is replayed faithfully, but the order of the episodes in the input stream is randomized. Repeated training iterations have an impact on the representation only because of the incremental nature of the SFA implementation we use.

2.3.2 | Generating novel sequences from episode fragments

The episodic system is represented by a highly simplified algorithmic model for sequence storage and retrieval, which was inspired by iterative retrieval from a hetero-associative network. In this model, each episode element or pattern y_i is stored individually. The sequential information is preserved by storing a retrieval key y^*_{i+1} pointing to the next pattern in the episode. Hence, the information about sequential order is only available on a pairwise basis (y_i, y^*_{i+1}), not on a global level.

Because the last element of an episode does not have a succeeding element that could be used as a key, it is not stored as a pattern explicitly. It is only stored as a key associated with the second-to-last episode element (Figures 4 and 5).

Retrieval of a sequence is initiated by providing a retrieval cue $\hat{y}_{t=0}$ to the system. The algorithm calculates the Euclidean distance of $\hat{y}_t$ to all patterns in memory and retrieves the one with the smallest distance, y'_i. The key y^*_{i+1} associated with y'_i is then used as a cue for the next retrieval step: $\hat{y}_{t+1} = y^*_{i+1}$. The process described so far is able to retrieve episodes from memory perfectly (except for the last pattern of an episode which cannot be retrieved, see above), unless two or more patterns from different episodes are identical. Sequence retrieval in biological neural networks, however, is subject to internal and external noise, we therefore add a noise term $\epsilon_t \sim N(0, \sigma)$, σ = 0.2 to the cue in each retrieval step (Equation 5). To avoid getting stuck in short loops during retrieval, a depression term is introduced: Every pattern y_i in memory is associated with a depression value a_i that is added to the distance to the cue during retrieval. a_i is initialized with 0 and is increased by a fixed amount α every time y_i is retrieved. Depression values decay exponentially with a decay constant of $\frac{1}{b}$. Equation (7) defines the depression term, with $\hat{u}_{i(t+1)}$ being a unit vector, in which the element at position i_{t+1} is 1.

$$d_t = (\|\hat{y_t} + \epsilon_t - y_0\|, \|\hat{y_t} + \epsilon_t - y_1\|, \ldots, \|\hat{y_t} + \epsilon_t - y_n\|) \text{ (Euclidean distances)} \quad (5)$$

$$i_{t+1} = \text{argmin}(d_t + a_t) \text{ (index of pattern to retrieve)} \quad (6)$$

$$a_{t+1} = a_t \cdot e^{-\frac{1}{b}} + \alpha \cdot \hat{u}_{i(t+1)} \text{ (depression term)} \quad (7)$$

In our simulations α and b are both set to 400. This provides enough immediate depression to avoid short retrieval loops, but the decay still allows one pattern to be retrieved multiple times during one simulation.

Because the last element of each episode is stored in memory only as a key (Figure 5, empty rings), it cannot be retrieved from memory. Therefore, when cued with one such pattern, the algorithm will retrieve a pattern that is similar to the cue, thus continuing the sequence in a smooth manner where the input stream normally would have exhibited a jump. These transitions are more frequent when the stored episodes are shorter. Figure 5 visualizes retrieval from episodic memory, contrasting episodes of length three and six. The loss of information by not storing the last pattern of each episode is negligible because in an episode consecutive patterns are similar. Furthermore, other episodes most probably contain similar patterns as well because of the high total number of patterns in memory (30,000 in our simulations, see below).

Taken together, retrieved sequences differ from original episodes due to two mechanisms: (a) Random retrieval errors. Their frequency depends on the level σ of retrieval noise. (b) Tying together episode fragments from memory that fit smoothly. The frequency of this happening depends on the number of episodes in memory.

We study the influence of episodic memory for episodes of different lengths, while keeping the total number of frames in the training data constant at 30,000. The number and length of episodes in the training data varies from 15,000 episodes of length two up to 50 episodes of length 600. From episodic memory, however, we always retrieve 375 sequences of length 80, each of which may be composed of fragments of several stored episodes.

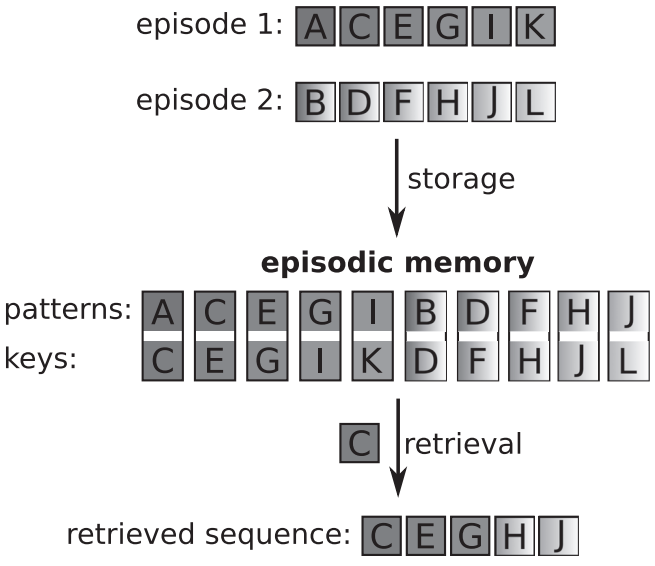

FIGURE 4 Illustration of the simplified model of episodic memory. The two episodes at the top are stored in the model of episodic memory as a set of pattern–key pairs. A key to the next pattern is stored along with each pattern. The last pattern of each episode (K, L) is not stored explicitly, that is, it only appears in memory as a key. During retrieval, noise is added to the retrieval cue and the closest memory pattern (Euclidean distance) to the noisy cue is retrieved. The corresponding key is used as the next retrieval cue. Due to the noise, retrieval can yield an incorrect transition (G → H in this example)

2.4 | Comparison of feature quality

After training SFA_{hi} with and without episodic memory, we assess the quality of their features using the following two different criteria.

1. The delta value (Equation 1) of the SFA_{hi} output on a set of testing data is used as a measurement of feature quality (e.g., Figure 8a).

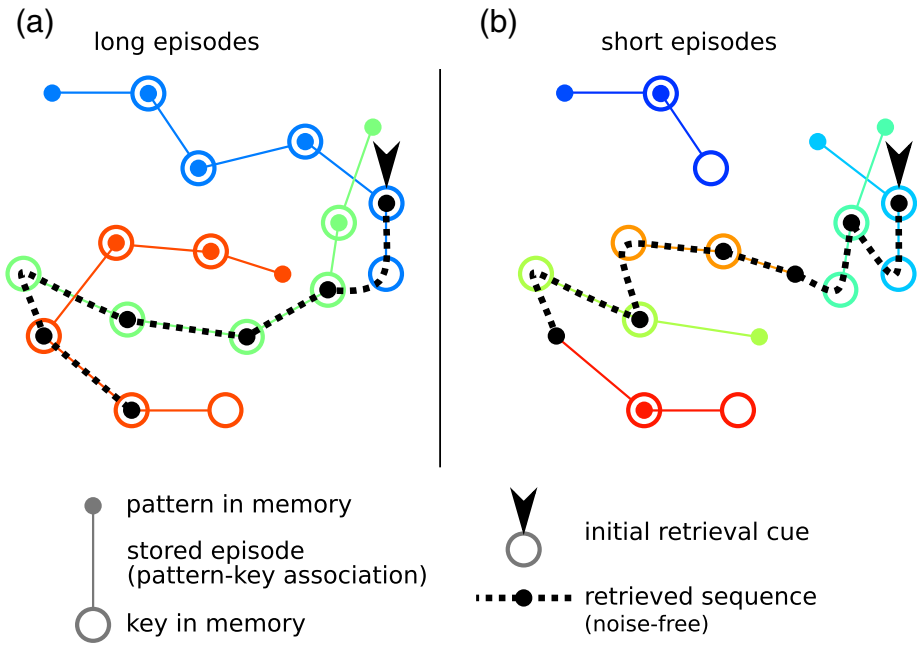

FIGURE 5 Visualization of noise-free sequence retrieval from episodic memory. Retrieval is compared for stored episodes of different lengths. Example patterns are visualized in two-dimensional space. Patterns and keys in memory are represented by filled circles and rings, respectively. The pattern–key associations stored in memory are depicted by solid lines. The initial retrieval cue is the key marked by an arrowhead. In each retrieval step the pattern closest to the cue is retrieved from memory. In the noise-free case this is the pattern identical to the key (the filled circle in the ring). However, if the end of a stored episode is reached, there is no pattern identical to the key in memory (circle empty) and the most similar pattern is retrieved (dashed line). After retrieving a pattern, the key associated (solid line) to that pattern is used as a cue for the next retrieval step. Dashed lines and black filled circles represent the retrieved sequence. (a) Episodic memory contains three episodes of length six. A sequence of length six is retrieved from memory. During retrieval, the end of a stored episode is reached twice. (b) Episodic memory contains six episodes of length three. A sequence of length six is retrieved from memory. During retrieval, the end of a stored episode is reached four times [Color figure can be viewed at wileyonlinelibrary.com]

Since the delta value is the objective function of SFA, features with a lower delta value are better in terms of the algorithm. To enable a meaningful comparison, SFA_{hi} output is whitened before calculating delta values. Note that the delta value does not evaluate the nature of the features. It quantifies the invariance of the representation, grounded in the assumption that a meaningful representation of information about a continuous input stream varies slowly in time.

2. We assess how well the latent variables are represented by the SFA_{hi} features. While the three slowest SFA_{hi} features could be made to code for the three latent variables (*x*-coordinate, *y*-coordinate, object identity) by separating the timescales on which the latent variables vary, this scenario is probably not a good description of most latent variables that humans encounter. We therefore did not design the input sequences that way and as a consequence there is no one-to-one relationship between features and latent variables, for example, one feature may code for a linear combination of *x*- and *y*-coordinate and react more strongly for one object than for the other. Hence, we assess how well linear combinations of the features correlate with each latent variable. To do so, we train a multivariate linear regressor to predict the latent variables given the first three SFA_{hi} features, using the training data.

2.5 | Visual discrimination performance

We follow the approach from one of the first studies that showed perceptual deficits in patients with MTL damage (A. C. H. Lee, Fischer, et al., 2005). Patients and controls viewed pairs of test images, along with a sample image to compare to. The task was to decide which one of the two test images was more similar to the sample image. The test image pairs were linear combinations of the shown sample image and a second sample image, which was not shown.

To simulate such a visual discrimination task, we use two different paradigms for image generation, each of which takes into account one of the latent variable types: Paradigm 1 mixes the sample images based on the coordinate of the object, Paradigm 2 mixes based on object identity, which models the original task more closely. By varying the level of similarity of the mixtures ("mixing level"), the difficulty of the task can be controlled for.

2.5.1 | Paradigm 1 (position discrimination)

The two sample images (S1/S2) show the letter T in different spatial positions. The distance between these two positions is fixed at 1 unit.

Based on these samples, test images (T1/T2) are generated by using different mixing levels. For example, for a mixing level of 0, T1 is identical to S1 and T2 is identical to S2. When the mixing level is increased, the letters in T1 and T2 are moved along a line toward each other until for a mixing level of 0.5, T1 and T2 are identical. An example set of image pairs is shown in Figure 6a.

2.5.2 | Paradigm 2 (object discrimination)

Two noise-free sample images, one showing the letter T at a random position (S1) and the other showing the letter L at the same position (S2), are generated. Based on these samples, test images (T1/T2) for different levels of mixing level m are generated. Test images are a linear combination of the two sample images:

$$\begin{aligned} T_1 &= (1-m)\cdot S_1 + m\cdot S_2 \\ T_2 &= m\cdot S_1 + (1-m)\cdot S_2 \end{aligned} \quad (8)$$

Gaussian noise with zero mean and a *SD* of 0.1 is added to the pixels of the images after mixing (gray values range from 0 to 1). An example set of image pairs according to Paradigm 2 is shown in Figure 6b.

In both paradigms, the task is to decide, which image, T1 or T2, is more similar to S1. T1 is always the correct answer. The images are first processed by SFA_{lo} and then by either SFA_{hi}[S] or SFA_{hi}[E]. In both cases, the resulting representations of T1 and T2 are compared to the representation of S1. If the Euclidean distance is smaller for T1 than for T2, the algorithm makes a correct decision in favor of T1. We evaluate 20,000 trials for each mixing level (0, 0.1, 0.2, 0.3, 0.4, 0.45, 0.475, 0.49, 0.5) and calculate the hit rate (percentage of correct decisions) for each of them.

3 | RESULTS

3.1 | The effect of memory on visual discrimination performance

When training the sensory representation, we used two different simplified memory models (Section 2.3). (a) Repeated verbatim replay of the episodes. The performance of the SFA_{hi}[E] module, trained on up to 40 repetitions of the training episodes, is compared to the performance of the SFA_{hi}[S] module that was exposed to the training episodes only once and serves as our model of MTL lesion. (b) Generating novel sequences from training episodes with a simple hetero-associative sequence storage model. Here SFA_{hi}[E] was trained on the newly generated sequences, while SFA_{hi}[S] was trained on the original episodes, again serving as our lesion model. Here the total amount of training data is identical between [E] and [S].

Additionally, we studied the influence of noise in the sequence storage model on the sensory representations. We compared the performance of the SFA_{hi}[E] module trained with noise-free memory ($\sigma = 0$) to the one trained with elevated noise levels in memory ($\sigma = 4$). Hence, the memory in this case is not completely absent, but retrieval is less reliable. This might be a more realistic model for the effect of partial lesions.

For all memory models, the discrimination performance is better with intact episodic memory in both discrimination paradigms (position as well as object discrimination, see Section 2.5; Figure 7),

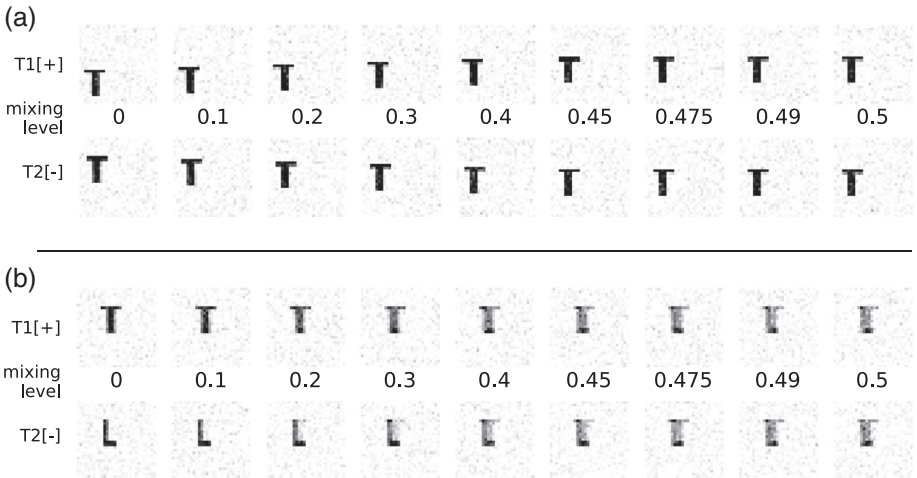

FIGURE 6 Sample stimuli for the simulated visual discrimination. (a) Example stimulus set for Paradigm 1. Two sample images (S1/S2) that only differ in the position of the letter T are generated (left-most images). The distance between the letter positions is fixed. The discrimination task is to decide which one of two different mixtures T1/T2 (shown as vertical pairs) is more similar to S1. The mixing level indicates the difficulty of the task. For mixing level 0.5 both T1 and T2 are identical (apart from noise). Reducing mixing levels increases the distance between T1 and T2 until, for mixing level 0, T1 = S1 and T2 = S2. (b) Example stimulus set for Paradigm 2. Two noise-free images are generated, one showing T at a random position and the other showing the letter L at the same position. Different linear mixtures of the two images are generated. Noise is added after mixing

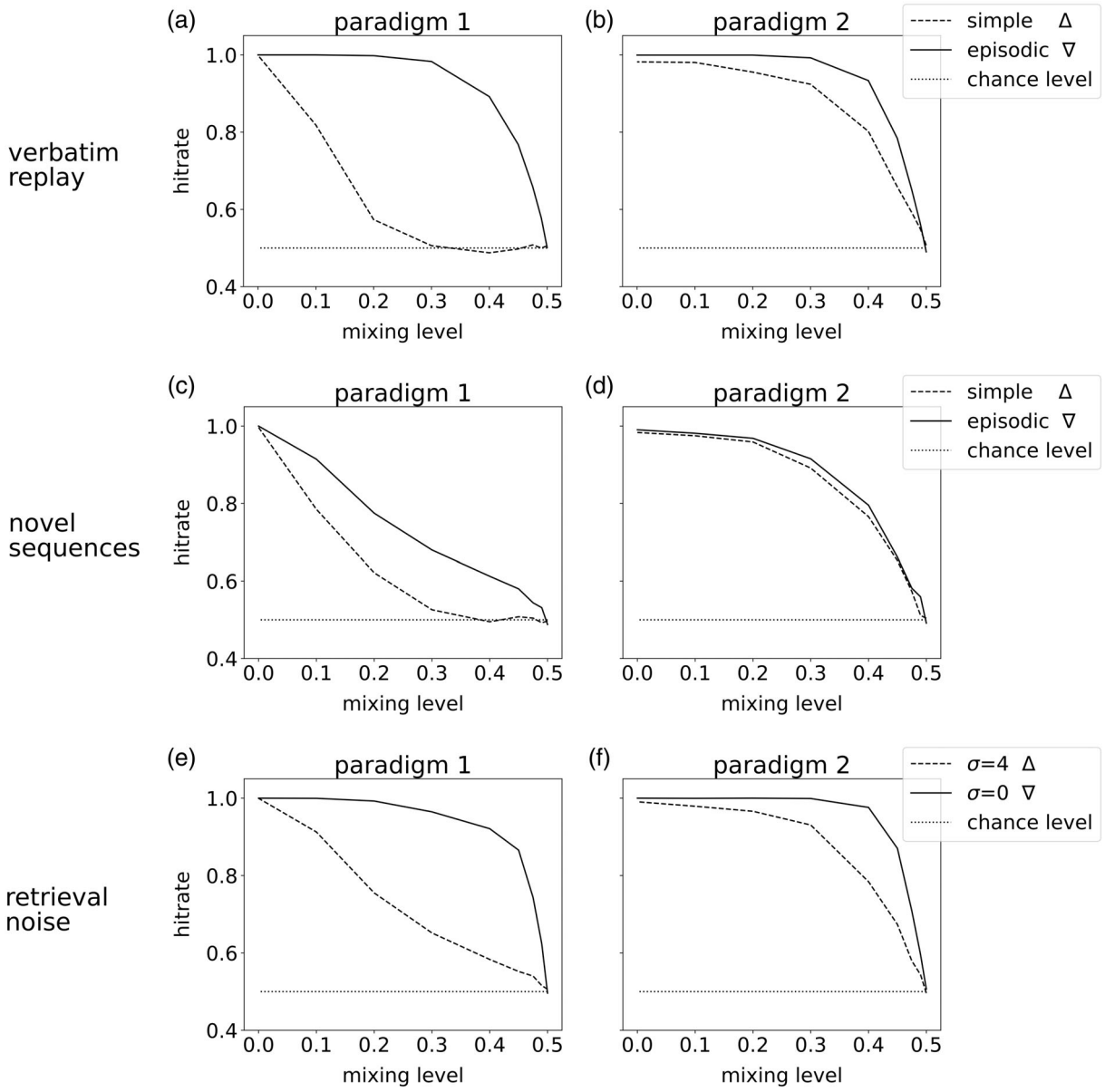

FIGURE 7 Sensory discrimination performance is superior with intact episodic memory. The dashed lines represent the hit rate of the sensory representation trained with lesioned memory, the solid lines represent performance with intact memory. The mixing level represents the level of similarity of the images to discriminate, which amounts to the level of difficulty of the discrimination. Each row of the figure shows the results from a different memory model as follows: (a and b) Repeated verbatim replay of the episodes. (c and d) Generating novel sequences from a hetero-associative sequence storage model. (e and f) Elevated versus no retrieval noise (σ) in the sequence storage model. The columns represent the two discrimination paradigms: (a, c, and e) Results for Paradigm 1 (position discrimination). (b, d, and f) Results for Paradigm 2 (object discrimination)

consistent with the finding that hippocampal patients are impaired in sensory discrimination tasks. Our results show that episodic memory can have a positive influence on the learning of sensory representations, not only by allowing to repeat the learning process (repeated verbatim replay) but also by providing memories in a more useful order (generating novel sequences). Our results suggest that an impairment in a visual discrimination task due to an MTL lesion need not imply an involvement of the MTL in visual processing. In our model, the influence of the MTL is indirect, through providing the memory for tuning the sensory representations. This is the main result of our study. In the following, we examine each memory model and the properties of the resulting sensory representations in detail to understand what difference the use of episodic memory exactly makes.

This analysis will eventually allow us to account for the performance difference between intact and lesioned memory, which, first, is more pronounced for simple replay than for generated novel sequences and, second, more pronounced in Paradigm 1 than in

Paradigm 2. Anticipating the results of our analysis below, the first effect is due to the fact that, in our simplified model, verbatim replay improves learning more than the generation of novel sequences. This might explain why replay is frequently observed in the hippocampus (Diba & Buzsáki, 2007; Louie & Wilson, 2001) and why it is important for learning (Girardeau et al., 2009).

The second effect arises because two different properties of the sensory representations are responsible for the differences in performance. In Paradigm 1, the difference arises mainly because the features in SFA_{hi}[E] code for the object coordinate much more precisely than SFA_{hi}[S] does. As a result, a distance in coordinate space yields a large distance in the feature space of SFA_{hi}[E] with a higher probability than in the feature space of SFA_{hi}[S]. In Paradigm 2, the performance differences only arise because the sensory representation of feature space of SFA_{hi}[E] is more robust to noise in the input than the feature space of SFA_{hi}[S] is. If no input noise was added, or the noise was added before calculating the linear combination of the sample images, using either SFA_{hi}[E] or SFA_{hi}[S] would yield perfect performance. This is the direct consequence of using only linear functions in SFA in our model. Since the output of SFA on a linear combination of two inputs equals the linear combination of the output of SFA on the individual inputs, the discrimination algorithm would be able to tell which mixture is closer to the reference image no matter how good, or bad, the representation of the input was. For more biologically plausible stimuli, nonlinear functions would have to be used in SFA, and we would expect larger differences in Paradigm 2. Hence, despite the performance difference between Paradigm 1 and 2 in this dataset, the model does not predict in general that position discrimination would profit more from an intact episodic memory than object discrimination.

3.2 | Evaluation of learned feature representations

3.2.1 | Repeated replay of the episodes

Replaying episodes faithfully from memory repeatedly is possibly the simplest form of sequential memory recall. It provides the neocortex with overall more training data and more opportunities to learn from the experienced episodes. Thus, we expect better sensory representations after multiple repetitions of memory replay. This is reminiscent of a process of systems consolidation, in which information from hippocampal memories is gradually extracted into neocortical memory stores with repeated retrievals.

Indeed, we find that the more often the training data are replayed, the better the SFA_{hi} features are in terms of their delta value (Figure 8a) and the better they represent the latent variables (Figure 8b–h) in the test data. While the time course of the delta value and the feature-latent-variable correlations are not identical, they are quite similar and asymptote by 40 presentations.

For the discrimination experiment we used two different instances of SFA_{hi}: The simple SFA_{hi}[S], which was trained only once on the training data and therefore serves as our model of sensory representation after hippocampal lesion, and the episodic SFA_{hi}[E], which was trained 40 times on the same data and serves as our model of normal sensory representation trained with episodic memory. Scatter plots show clearly that the features extracted by SFA_{hi}[E] are much better correlated with the latent variables (Figure 8f–h) than those of SFA_{hi}[S] (Figure 8c–e). These results show that episodic memory can enable the learning of an adequate sensory representation when only a limited amount of experience is available, by simple replay of those.

For comparison, we trained another SFA_{hi} instance by generating 40 unique data sets, instead of using the same data set 40 times. This new network was thus trained on the same overall amount of training data, but it has been exposed to more unique sensory stimuli than the network trained by replaying a limited amount of experience from memory. Interestingly, the replay network outperforms the network trained on more data (Figure 8a,b). This additionally emphasizes the importance of memory replay for learning.

In order to visualize the linear transformation that SFA_{hi}[E] learned, which we call the sensory representation, we plotted the response of the entire network to stimuli at different positions (Figure 9). As stimuli we used a single black pixel (Figure 9b,d) and noisy images containing the letters T (Figure 9c,e; top row) and L (Figure 9c,e; bottom row). The responses to the single black pixel were normalized by subtracting the answer to a uniform white stimulus. The figure shows that the Features 1 and 2 of SFA_{hi}[E] are sensitive to the position of the black pixels and exhibit a spatial gradient (Figure 9b, c). Thus, Features 1 and 2 mainly extract the *x*- and *y*-coordinates of the object in the input by applying a weight gradient. Feature 3 appears to be sensitive more to the number of black pixels (which differs between L and T), while being mostly invariant to their position. Hence, Feature 3 mainly extracts the identity of the object in the input by simply counting the number of black pixels. These properties are even more pronounced in the output of the linear regressor (Figure 9d,e).

3.2.2 | Generating novel sequences by tying together episode fragments

Episodic memory can also be used to improve training performance without increasing the total amount of data. When modeled with a hetero-associative sequence storage model (as described above, Section 2.3), the hippocampal system can associate similar patterns with each other that were experienced at different times. It has been shown that this sequential or associative property of the hippocampal memory is exploited for inference or generalization (Bunsey & Eichenbaum, 1996; Wimmer & Shohamy, 2012; Zeithamova, Dominick, & Preston, 2012). We hypothesize that these properties can also be used in our case of training the sensory representations. By associating similar episode fragments with each other, novel sequences can be generated from memory that are more useful to the representational learning system than the original episodes.

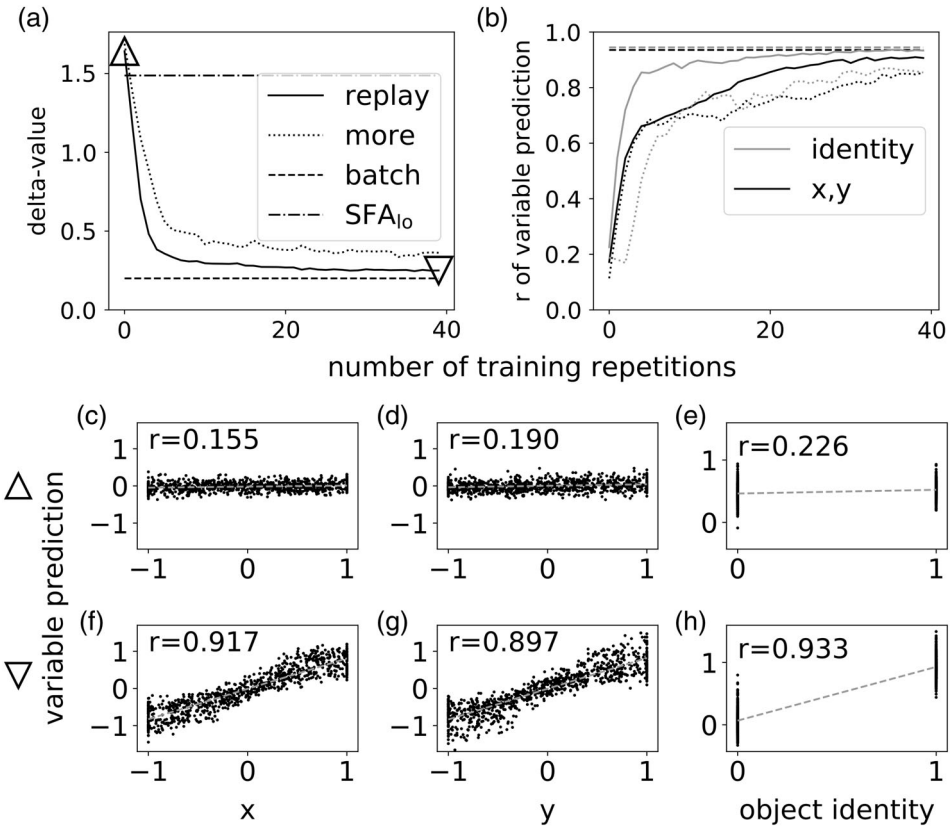

FIGURE 8 Repeated replay of episodes from memory improves sensory representations. Training was conducted 40 times, shuffling the order in which the episodes are presented in each repetition. The quality of the extracted features was evaluated on test data after each repetition. (a) Average delta value of the three slowest features. For comparison, the plot also shows the average delta value if new training data are generated in each repetition instead of replaying from memory ("more"). Besides, delta values are shown for the pretrained SFA_{lo} and for a batch SFA_{hi} using the 12-fold amount of training data ("batch"). Triangles denote the SFA_{hi} instances that were used in (c)–(h) and for the visual discrimination task. (b) Correlation of SFA_{hi} feature output with latent variables in the input. A multivariate linear regressor, for *x*, *y*, and object identity, was trained on the first three features using the training data. The regressor was then used to reconstruct the latent variables from SFA_{hi} feature output on the test data. The figure shows Pearson correlations (*r*-values) between original and reconstructed coordinates (c–h). *r*-Values of the *x*- and *y*-coordinates are averaged. (c–h) Predictions of regressor versus latent variables. A subsample of 1,000 data points is shown. The dashed line represents a linear regression. (c)–(e) show the data points for SFA features after the first training, (f)–(h) show data after all 40 training iterations. SFA, slow feature analysis

In our model, the two different objects are presented alternately, that is, at the start of a new episode the object identity is switched. Also, the position of the object is randomly re-initialized. When the episodes are strung together to form the training data set, the latent variables (object coordinates, identity) jump at the transitions between episodes (see also Section 2.1; Figure 2). These discontinuities make it harder for SFA to learn a representation of the latent variables, because SFA learns features that vary slowly in time. The shorter the episodes in the training data are, the more discontinuities with regard to the latent variables are present in the data (given that the overall amount of data is constant). This is reflected in the delta values (average rate of change) of the latent variables in the training data (Figure 10c, dashed lines). Hence, we expect the quality of the sensory representation learned to be lower with shorter episodes.

The hetero-associative sequence storage model of episodic memory, however, counteracts the discontinuities in the original data by associating similar episodes or episode fragments with each other, generating smooth sequences of fixed length. A data set composed of these sequences has a fixed number of discontinuities, which is reflected in the delta values (Figure 10c, solid lines). Hence, if the sequences retrieved from episodic memory are used to train SFA we would expect the sensory representation to be independent of the episode length in the training data.

We trained SFA_{hi} instances on episodes of different lengths between 2 and 600 while keeping the overall amount of data constant. SFAhi[S] is trained on these episodes directly, whereas SFAhi [E] is trained on sequences retrieved from episodic memory. This process is averaged over 16 repetitions to smooth out fluctuations that are introduced by randomness in the movement statistics of the input data, the selection of initial retrieval cues and the retrieval noise in episodic memory.

As expected, the longer the episodes in the training data are, the more precisely the features generated by SFA_{hi}[S] represent the latent variables of the input, that is, delta values decline (Figure 10a) and

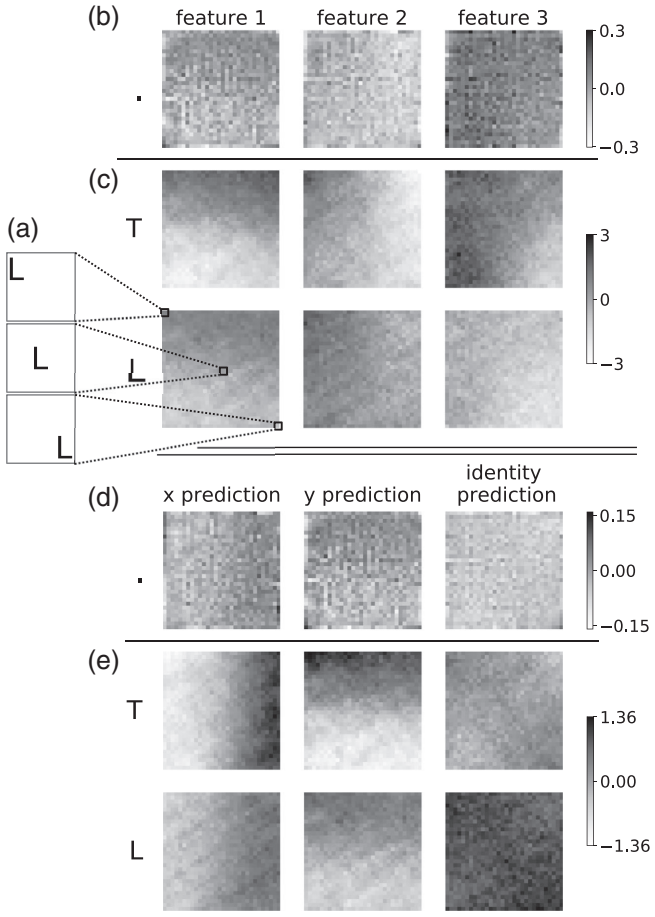

FIGURE 9 Sensory representations of SFA_{hi}[E] for stimuli at different positions. SFA_{hi} training was conducted 40 times. The value of each pixel is the response to an object centered on the respective position of the input image. (a) Legend depicting how to read the plots (b)–(e). The legend shows three example input images and illustrates which pixels of the plot represent the responses to these images. The top left pixel of the plots, for instance, is the response of the system to an object in the top left corner of the input image. Hence, if the pixel values of a plot display a gradient, that means the system responds differently to objects at different positions—the feature "codes" for object position. (b) Stimuli were single black pixels. Data were normalized by subtracting the representation of a zero image. From left to right: Features 1, 2, 3. (c) Stimuli were noisy images with either the letter T (top row) or L (bottom row). From left to right: Features 1, 2, 3. All plots in (b) and (c) display a spatial gradient, especially feature 1 and 2 code for object position. Feature 3 additionally displays a clear response difference between objects T and L, hence coding for object identity. (d and e) The same information as in (b) and (c) is depicted for the predictions of the linear regressor for *x*-coordinate, *y*-coordinate, and object identity (from left to right). Response gradients for the letters in Column 1 and 2 (*x*,*y*) and object specificity for Column 3 (identity) are more pronounced than in (b). This was expected since the SFA features are learned in an unsupervised manner, whereas the linear regressors were trained using ground truth as supervisory signal. SFA, slow feature analysis

feature-latent-variable correlations increase (Figure 10b). By contrast, when episodic memory is used to generate the training data for SFA_{hi}[E], the feature quality is almost independent of the length of the original episodes and is generally higher, but especially so for short episodes. Note that testing was always performed on sequences of length 50, regardless of how the sensory representations were trained, to facilitate a fair comparison across different training conditions (with/without episodic memory, different lengths of episodes).

This result shows that episodic memory can improve the sensory representations by providing training data in a more useful order. While this main effect is only substantial for shorter episodes in our model, it is conceivable that it is more general for real sensory inputs, which are much more complex. In order to optimize representations for different aspects of sensory percepts, episodic memory content could be reorganized according to different criteria. Notably, the identity of the object is harder to extract for the SFA algorithm than the position, thus optimally representing the object identity in the feature output requires longer training episodes than representing *x*- and *y*-coordinate (Figure 10b). Gupta et al. (2010) recorded place cell sequences from rats that the animals never experienced. They suggest that this observation reflects mechanisms to learn a complete representation of the environment. Recombining episode fragments could have a similar function in tuning sensory representations.

For the visual discrimination we used SFA_{hi}[S] and SFA_{hi}[E] instances from the trials where training episodes were of length 2, because the strongest effect of episodic memory was observed in that case.

Aside

In addition to the main effect described above, the data show two other, more nuanced, effects. Even though they do not affect our main results in any way, we explore the origin of these effects in the following to precisely understand the behavior of the model. This section can be skipped on a first reading.

1. Even for the longest training episodes, feature quality is higher in the episodic scenario, at least when measured by delta value of the SFA output. If the reason for the performance advantage of SFA_{hi}[E] over SFA_{hi}[S] was only that it is trained on data containing fewer jumps, one would expect that SFA_{hi}[S] performs better than SFA_{hi}[E] for episodes of a length larger than 80, which is the length of sequences generated by episodic memory. Since episodic memory is cued randomly, it will repeat at least parts of the episodes multiple times with a high probability. Because there are as many unique patterns as there are patterns in the training set, this repetition of some patterns implies that other patterns are not used at all, thus there are fewer unique patterns presented in the episodic than in the simple scenario. We show above (Section 3.2.1) that the system profits more from a repetition of input data than from being trained on more unique patterns. This explains the described effect – in the episodic scenario the training data contains repetitions, while in the simple scenario every input pattern is unique.
2. Counterintuitively, the delta value of the SFA_{hi} features increases slightly and the precision of their object coordinate representation

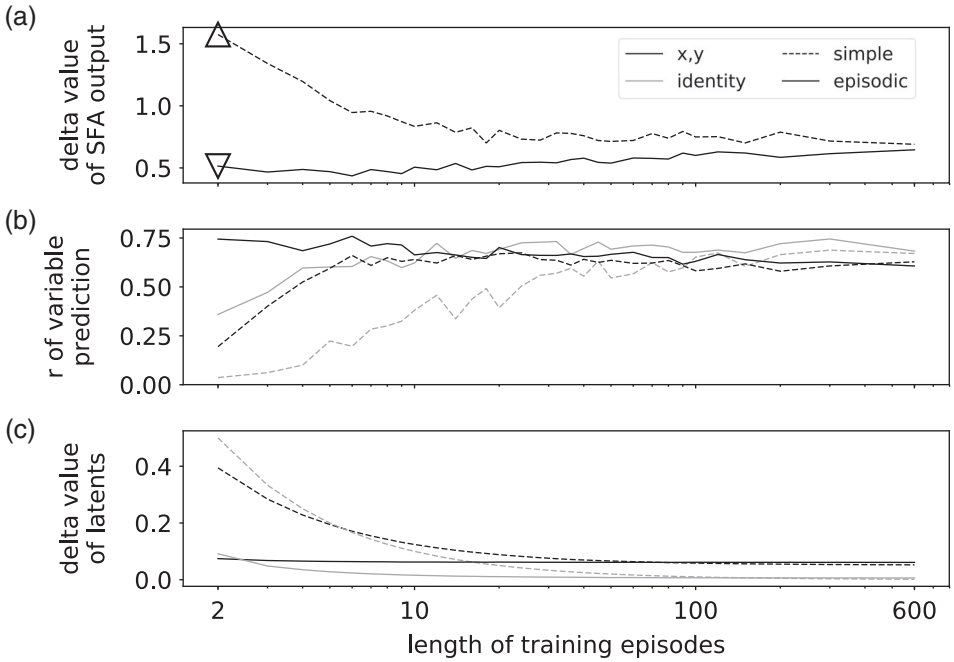

FIGURE 10 Generating novel sequences using episodic memory improves sensory representations. The overall amount of training data was fixed at 30,000 frames, but the number and length of individual episodes varied from 15,000 episodes of length two up to 50 episodes of length 600. From these episodes, episodic memory always generated 375 sequences of length 80. We contrast the feature quality between the simple and the episodic scenario, depending on the length of the training episodes. In the simple scenario, those episodes were used for training SFA_{hi}[S] directly, whereas in the episodic scenario, they were stored in episodic memory first, which then generated sequences with fixed length to train SFA_{hi}[E] on. (a) Average of the delta values for the three slowest features on the test data. Triangular markers in the plot denote the SFA_{hi} instances that were used for the visual discrimination. (b) Correlation of SFA_{hi} feature output with latent variables of the input on test data. As in Figure 8, this is the Pearson correlation between latent variables and estimates by the regressors. (c) Average delta value of the latent variables, that is, coordinates and binary object category in the training data. In the episodic scenario, this is the output of episodic memory. SFA, slow feature analysis

decreases slightly with increasing training episode length in the episodic scenario (Figure 10a,b). We found that this effect arises because of the property described above as well: With 30,000 training patterns at hand, SFA_{hi} performs better when some of the patterns are left out while others are repeated during training instead of presenting the full data set. Because the last pattern of each episode cannot be retrieved from memory, there are actually only 15,000 patterns in memory for training episodes of length 2. On the contrary, for an episode length of 600 a total of 29,950 patterns are stored in memory. As a consequence, the probability of retrieving the same patterns multiple times is lower for longer episodes, thus feature quality of SFA_{hi}[E] decreases slightly with increasing episode length. We tested this hypothesis by choosing depression parameters (Equation 7) such that a pattern would not be retrieved multiple times and by storing a sufficiently large number of patterns in memory, so that 375 smooth sequences of length 80 could be retrieved. When SFA_{hi}[E] was trained on those sequences, the delta value of the features is independent of the episode length (results not shown), confirming our hypothesis.

These two effects further emphasize that replaying the same episodes causes a larger improvement of the sensory representations than providing the model with more unique episodes, a result we found above (Section 3.2.1; Figure 8a,b).

3.2.3 | Noise in episodic memory

The above approaches have compared a memory-free model to a model with a fully functioning memory system. In reality, when the MTL is damaged or in the case of age-related impairment, episodic memory might not be completely absent. Also, memory models of forgetting in the healthy suggest that memory is not completely lost, but that retrieval can fail or be incorrect due to interference of memory traces (Anderson, 2000; Underwood, 1957) or inaccurate retrieval cues (Tulving, 1974). We model these various effects with increased levels of retrieval noise in our hetero-associative memory. Gradually increasing the retrieval noise in the hetero-associative sequence storage model will lead to associations of more and more dissimilar patterns, up to the point at which retrieval is completely random. All other aspects of the model remain unchanged. As before, we generated 30,000 frames of training data that are stored in episodic memory. While SFA_{hi}[S] was trained on the data directly, training of SFA_{hi}[E] used sequences retrieved from memory that consist of a total number of 30,000 frames as well. The length of individual episodes in

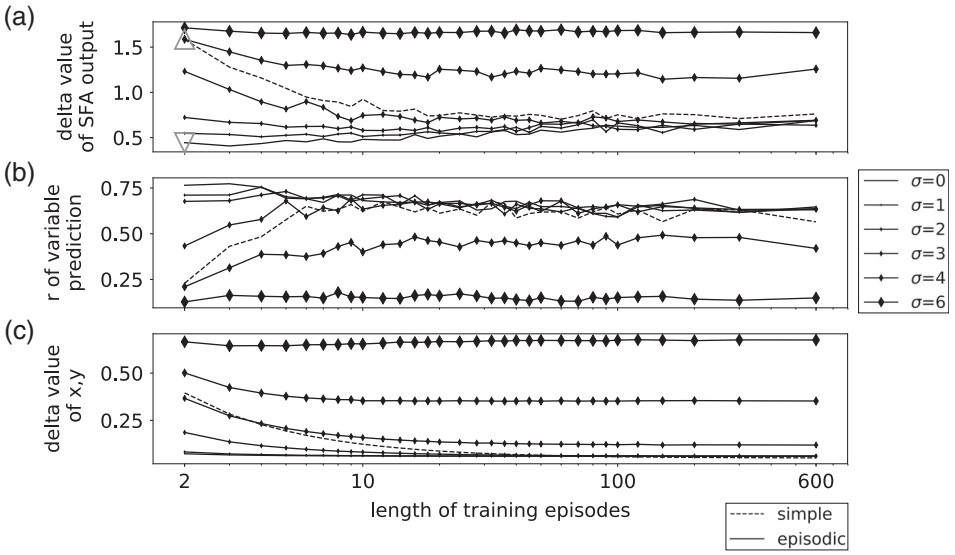

FIGURE 11 Retrieval noise (σ) in episodic memory reduces the quality of features extracted by SFA_{hi}[E]. Larger line markers denote more noise. (a) Average of the delta values of the three slowest features on test data. Triangular markers in the plot denote the SFA_{hi} instances that were used for the visual discrimination. (b) Correlation of SFA_{hi} feature output with *x*- and *y*-coordinate of the input on test data. (c) Average delta value of latent variables, that is, coordinates, of the training data for SFA_{hi}. In the episodic scenario this is the output of episodic memory. Note that the lines for $\sigma = 0$ and $\sigma = 1$ are almost identical. For legibility, the results relating to object identity are not shown, but they display the same effect as the coordinates

the training data was varied (*x*-axis in Figure 11), whereas the sequences retrieved from episodic memory always had a length of 80 patterns. Note that training SFA_{hi}[E] with large noise in episodic memory is qualitatively different from training SFA_{hi}[S] (without episodic memory), hence varying the level of retrieval noise does not yield a gradual transition from the episodic to the simple scenario.

As expected, with higher retrieval noise the feature quality is generally lower, that is, the features vary quicker (higher delta values) and represent the latent variables of the input less precisely (lower feature-latent-variables correlations; Figure 11). Retrieval noise leads to jumps to incorrect elements that are farther away from the previous element than the correct next element is. As a result, the delta values in the latent variables of the retrieved sequences are higher for larger noise (Figure 11C).

For the visual discrimination we used SFA_{hi}[E] instances from trials with training episodes of length 2. Intact episodic memory was modeled by using a retrieval noise level of $\sigma = 0$ and impaired episodic memory was modeled by an intermediate level of retrieval noise $\sigma = 4$.

Aside

This section examines a theoretical detail of the model behavior and can be skipped on a first reading. Apart from the main effect—the reduction of feature quality with increased retrieval noise—the data exhibit a second feature. For intermediate levels of retrieval noise, ($2 \leq \sigma \leq 4$), feature quality increases with episode length, before reaching an asymptote at an episode length of 5–10 (in the episodic scenario), unlike for low ($\sigma < 2$) or high ($\sigma = 6$) noise where the curves are flat. This effect is a purely theoretical result and does not affect our aforementioned main result. However, since the cause of this effect is not immediately obvious it warrants more detailed consideration. There are two ways in which an element might be retrieved that is different from the key. A: The model retrieves the element closest to the cue, which is the sum of the key and retrieval noise (Equation 5). The larger the retrieval noise, the more likely it is that a jump to an incorrect item occurs and the larger the jump. This gives rise to the main effect discussed above. B: If the key refers to the last element of an episode, a different element is retrieved because the last element of an episode is not stored in our model. This kind of jump occurs whether there is retrieval noise or not (see Section 3.2.2). The second effect discussed here is a result of combining A and B. The shorter the episodes, the more often the end of an episode is reached. When this happens in the presence of little or no noise ($\sigma < 2$), the element closest to the key is retrieved, leading to almost no change in delta value of the underlying latent variables (Figure 11C). In the presence of intermediate noise, the closest element to the cue = key + noise is in many cases different from the element closest to the key and therefore further away from the previous element (Figure 12). This leads to a bigger jump at the end of an episode, which does not occur in the middle of an episode, because in the latter case, the correct element is stored in the system and will be retrieved correctly, as long as the noise is not too large. Thus, these bigger jumps occur more frequently the shorter the episodes in memory are, which is reflected in the delta values of the latent variables (Figure 11C) and the resulting feature quality (Figure 11a,b). For very large retrieval noise ($\sigma = 6$), the noise vector is larger than the average inter-item distance and so it makes no difference whether the element associated with the key is stored in the system or not (Figure 12). Hence, for large noise the quality of the extracted features does not depend on the length of the stored episodes (Figure 11).

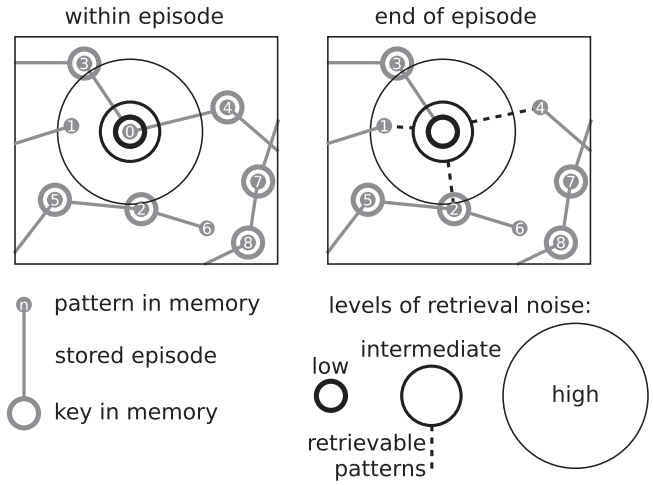

FIGURE 12 Aside: intermediate levels of retrieval noise selectively increase pattern distances at the end of episodes. Patterns and keys in memory are represented by gray filled circles and rings, respectively. The pattern–key associations stored in memory are depicted by gray lines. Suppose pattern No. 3 was just retrieved. The next retrieval step will be cued by the key associated with pattern No. 3 (thick black ring). *Left*: Retrieval step within an episode. The pattern (No. 0) corresponding to the key (thick black ring) is available in memory. *Right*: Retrieval step within an episode. The pattern corresponding to the key is not available in memory (pattern No. 0 is missing in the thick black ring). Different levels of retrieval noise are depicted by black rings with different radii (low, intermediate, high). After applying noise, the noisy retrieval cue can be anywhere within the respective ring. Retrieval always proceeds by retrieving the pattern closest to the noisy cue. If the noise is low, the noisy cue will not deviate much from the stored key, hence the retrieved pattern will be always the one closest to the key (left panel: No. 0, right panel: No. 1). When the noise is increased, the variability of the noisy cue is at some point so high that a number of different patterns can be retrieved, that have a larger distance to the previous pattern. For the retrieval steps within an episode (left panel), this point is reached only for high levels of noise, but for the retrieval steps at the end of an episode (right panel), this happens already at intermediate noise levels. This differential effect is strongest for short episodes, when there are more episode endings. Thus, feature quality is impaired more by intermediate noise for short episodes than for long episodes

4 | DISCUSSION

We have proposed an algorithmic model for how episodic memory in MTL can improve sensory discrimination performance by helping to optimize the representation of sensory information in neocortex. We separately investigated the effects that episodic memory has when it (a) repeatedly replays the memory of the original episodes faithfully or (b) generates novel sequences by tying together episode fragments from memory. We found that, in our model, 40-fold replay from episodic memory (total of 200,000 patterns) leads to a substantially better sensory representation. The generation of novel sequences from episodic memory (total of 30,000 patterns) can also have a beneficial effect when the individual original episodes are short and the retrieval noise is low. Most importantly, we found that an optimized sensory representation can be advantageous for perceptual decision making. In simulated visual discrimination tasks, the systems that were trained using episodic memory outperformed the systems without episodic memory.

While our results were obtained with a particular kind of sensory representation extracted using SFA, there is a case to be made that our finding would extend to other types of models as well. Using episodic memory to replay the memory of the sensory experience faithfully is equivalent to repeated training iterations on the same data set. It is likely that replay benefits any incremental learning algorithm, since many machine learning algorithms and biological agents require multiple repetitions to converge to an optimum, when extracting information from training data. For instance, reinforcement learning algorithms profit greatly from experience replay (Lin, 1993; Mnih et al., 2015). Hence, our findings that replay helps improve sensory discrimination is likely not limited to the particular algorithm that we have used here.

The benefit of retrieving sequences composed of episode fragments from memory might not be quite as general. The generation of novel sequences is equivalent to changing the order of presentation of individual samples. This transposition only makes a difference for learning algorithms that are sensitive to the temporal order in the training data. Such learning algorithms are probably employed by biological systems that have to represent temporal and possibly causal relationships. For instance, it has been suggested that the hippocampus helps to establish associations between spatially and temporally discontiguous events (Pyka & Cheng, 2014; Wallenstein, Eichenbaum, & Hasselmo, 1998). In our model, this is reflected in the sequential nature of episodic memory and the generation of novel sequences during which temporally discontiguous patterns can be associated. The particular algorithm we used in this study, SFA, is just one example of an algorithm that is sensitive to temporal order. It was previously shown to be well-suited for modeling realistic neuronal responses in the visual system (Berkes & Wiskott, 2005) as well as the hippocampus (Franzius, Sprekeler, & Wiskott, 2007), and so appears to be a reasonable choice for an algorithmic model of sensory representations.

Similarly, the memory in our model need not necessarily be episodic memory. In our study, we are looking at how the MTL can have an influence on perceptual tasks, and since memory in the human MTL has been linked strongly to episodic memory, we only refer to episodic memory. However, other types of memory could have a similar effect on neocortical representations by improving the quantity or the quality of the training data. Our model, though, has features that are consistent with episodic memory, namely that it has sensory content and that it is sequential.

4.1 | Sensory representations for visual discrimination

Our results suggest that visual discrimination is more accurate when the sensory representation has been tuned by intact episodic memory. Although the stimuli used in the model are highly simplified for

computational reasons, the same principles that we have identified should apply to more complex images as well. Processing those more complex images would require a more advanced sensory representation, which would take longer to train and perhaps consist of more layers. Hence, it could probably profit even more from episodic memory than our very simple representation.

The perceptual task that we modeled was a discrimination task, which involves only two comparisons. However, the procedure to perform an oddity judgment task (Barense, Gaffan, & Graham, 2007; Lech & Suchan, 2014; A. C. H. Lee, Buckley, et al., 2005) based on sensory representations is similar, since odd-one-out judgments could be made on the basis of pairwise comparisons. The odd-one-out will then be determined as the one stimulus with the highest average difference to the other stimuli. Thus, the findings of lesion experiments based on visual discrimination or oddity judgment tasks could both be accounted for by the tuning of sensory representations driven by episodic memory.

In other perceptual tasks, even when they do not involve stimulus comparisons, the participant's decision still has to be based on a sensory representation that processes the features of the image or the visual scene. In a study by Lee et al. (A. C. H. Lee & Rudebeck, 2010), participants had to judge line drawings on their geometrical plausibility. A patient with MTL lesions showed impaired performance and different fixation patterns as compared to healthy controls. When the sensory representation is not well optimized—for example, to represent geometry and perspectivity of the stimulus—it is harder to detect regions of interest and possible geometrical flaws. Furthermore, it is not obvious whether this geometric plausibility judgment is a purely perceptual task, since it requires experience with, and analysis of, geometrical drawings.

4.2 | MTL activation during perceptual tasks

While in our model memory stored in MTL affects the properties of the sensory representation, which can account for the results of lesion studies, the MTL is not involved in the visual discrimination process itself. Our model thus does not predict the MTL activation that has been observed during certain perceptual tasks in fMRI studies (Barense, Henson, Lee, & Graham, 2010; Lech & Suchan, 2014; A. C. H. Lee, Bandelow, Schwarzbauer, Henson, & Graham, 2006; A. C. H. Lee, Scahill, & Graham, 2008). While the perceptual-mnemonic hypothesis attributes this activation to the direct involvement of the MTL in perception, incidental memory encoding processes could be an alternative explanation, especially because the presented stimuli are usually trial-unique (A. C. H. Lee et al., 2008). In some studies, the authors acknowledge this possibility while others are designed to control for that. For instance, Lech and Suchan (2014) controlled for incidental encoding by conducting an additional recognition task after the visual task and comparing the recognition of the studied items to the recognition of the items from an oddity judgment task. Since significantly fewer of the items from the oddity task were recognized as compared to the studied items, the authors concluded that no incidental encoding had occurred. However, although the recognition rate on the studied items (~50%) is indeed higher than on the items from the oddity task (~30%), a comparison to random performance is not possible because the false alarm rate was not given. Hence, it cannot be excluded that items from the oddity task have been stored and successfully retrieved from memory. Furthermore, it is not surprising that the memory of incidentally encoded items is weaker than that of explicitly encoded items, especially because the memory had been maintained over a delay period of 1 week.

In another fMRI study by Lee et al. the pool of initially trial-unique stimuli was used three times in an oddity judgment task in order to investigate whether MTL activation decreases when a stimulus is presented repeatedly (A. C. H. Lee et al., 2008). Indeed, the authors found clear evidence of incidental memory encoding in posterior hippocampus and parahippocampus during oddity judgments. For perirhinal cortex and anterior hippocampus the evidence is less clear, but incidental encoding could not be ruled out, especially because the participants' performance on the task improved across the three sessions.

4.3 | Functional subdivision of the MTL

Studies often suggest differential roles for the hippocampus and the perirhinal cortex in perception, depending on the stimulus material (Barense et al., 2010; A. C. H. Lee, Buckley, et al., 2005; A. C. H. Lee, Fischer, et al., 2005; A. C. H. Lee et al., 2008). It has been proposed that the hippocampus is involved in the perception of scenes, whereas the perirhinal cortex in involved in the perception of faces and complex objects. In lesion studies, performance is compared between patients with localized hippocampal damage and patients with extensive MTL damage, which indeed includes the perirhinal cortex, but also other structures (A. C. H. Lee, Fischer, et al., 2005; Shrager et al., 2006; Suzuki, 2009). Moreover, the hippocampus is usually damaged to a larger degree in MTL patients as compared to hippocampal patients. The influence of the perirhinal cortex on perception is then derived by subtracting the effect of hippocampal lesions from that of MTL lesions. It is possible that the resulting difference reflects the specific contribution of the perirhinal cortex to perception. Alternatively, the lesion size alone could account for the different impairments of hippocampal and MTL patients (Suzuki, 2009). According to this view, scene perception is already impaired by a small lesion, while it requires a larger lesion to impair the perception of objects and faces. Furthermore, the influence of lesion size on task performance might not be linear, such that a small increase in lesion size could have a large effect on visual perception, or vice versa.

Studies using fMRI are inconclusive in this regard. Some do report differential activation of the hippocampus and the perirhinal cortex depending on the type of stimulus (Barense et al., 2010; A. C. H. Lee et al., 2008), while others could not reproduce that finding (Lech & Suchan, 2014; A. C. H. Lee et al., 2006). This leaves room for speculation that the reported activation differences are not attributable to different stimulus categories, but to different stimulus complexities or differences in the low-level features, for example, round or sharp

edges, textures, and so on. For instance, it has been shown that second-order image statistics differ between image categories (Torralba & Oliva, 2003).

We conclude that the influence of the MTL on perceptual processes might stem from its mnemonic function, not from a direct role in the perceptual process. Increased activation in MTL areas during perceptual tasks might reflect task-irrelevant, memory-related activity. Patients with damage to the MTL might be impaired in perceptual tasks because their sensory representation is less optimized, due to the limited availability of episodic memory for the tuning of sensory representations.

ACKNOWLEDGMENTS

The authors thank Boris Suchan for helpful comments on the manuscript. This work was supported by grants from the German Research Foundation (DFG) through the SFB 874, project B2—project number 122679504 (S.C.) and through the FOR 2812, project P5—project number 419039588 (L.W.), and from the German Federal Ministry of Education and Research (BMBF), grant 01GQ1506 (S.C.).

DATA AVAILABILITY STATEMENT

The code for the computational model is openly available at https://doi.org/10.5281/zenodo.3578454

ORCID

Richard Görler https://orcid.org/0000-0002-6076-0875
Laurenz Wiskott https://orcid.org/0000-0001-6237-740X
Sen Cheng https://orcid.org/0000-0002-6719-8029

REFERENCES

Agster, K. L., Fortin, N. J., & Eichenbaum, H. (2002). The hippocampus and disambiguation of overlapping sequences. *The Journal of Neuroscience*, *22*(13), 5760–5768. https://doi.org/10.1523/JNEUROSCI.22-13-05760.2002

Anderson, J. R. (2000). *Learning and memory: An integrated approach* (2nd ed.). Hoboken, NJ: John Wiley & Sons Inc.

Barense, M. D., Gaffan, D., & Graham, K. S. (2007). The human medial temporal lobe processes online representations of complex objects. *Neuropsychologia*, *45*(13), 2963–2974. https://doi.org/10.1016/j.neuropsychologia.2007.05.023

Barense, M. D., Henson, R. N. A., Lee, A. C. H., & Graham, K. S. (2010). Medial temporal lobe activity during complex discrimination of faces, objects, and scenes: Effects of viewpoint. *Hippocampus*, *20*(3), 389–401. https://doi.org/10.1002/hipo.20641

Bayati, M., Neher, T., Melchior, J., Diba, K., Wiskott, L., & Cheng, S. (2018). Storage fidelity for sequence memory in the hippocampal circuit. *PLoS One*, *13*(10), e0204685. https://doi.org/10.1371/journal.pone.0204685

Berkes, P., & Wiskott, L. (2005). Slow feature analysis yields a rich repertoire of complex cell properties. *Journal of Vision*, *5*(6), 579–602. https://doi.org/10.1167/5.6.9

Beyeler, M., Rounds, E. L., Carlson, K. D., Dutt, N., & Krichmar, J. L. (2019). Neural correlates of sparse coding and dimensionality reduction. *PLOS Computational Biology*, *15*(6), e1006908. https://doi.org/10.1371/journal.pcbi.1006908

Buckley, M. J., Booth, M. C., Rolls, E. T., & Gaffan, D. (2001). Selective perceptual impairments after perirhinal cortex ablation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *21*(24), 9824–9836. https://doi.org/10.1523/JNEUROSCI.21-24-09824.2001

Buhry, L., Azizi, A. H., & Cheng, S. (2011). Reactivation, replay, and preplay: How it might all fit together. *Neural Plasticity*, *2011*, 1–11. https://doi.org/10.1155/2011/203462

Bunsey, M., & Eichenbaum, H. B. (1996). Conservation of hippocampal memory function in rats and humans. *Nature*, *379*(6562), 255–257. https://doi.org/10.1038/379255a0

Bussey, T. J., & Saksida, L. M. (2007). Memory, perception, and the ventral visual-perirhinal-hippocampal stream: Thinking outside of the boxes. *Hippocampus*, *17*(9), 898–908. https://doi.org/10.1002/hipo.20320

Cai, D. J., Aharoni, D., Shuman, T., Shobe, J., Biane, J., Song, W., … Silva, A. J. (2016). A shared neural ensemble links distinct contextual memories encoded close in time. *Nature*, *534*(7605), 115–118. https://doi.org/10.1038/nature17955

Cheng, S. (2013). The CRISP theory of hippocampal function in episodic memory. *Frontiers in Neural Circuits*, *7*, 88. https://doi.org/10.3389/fncir.2013.00088

Cheng, S. (2017). Consolidation of episodic memory: An epiphenomenon of semantic learning. In N. Axmacher & B. Rasch (Eds.), *Cognitive neuroscience of memory consolidation* (pp. 57–72). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-45066-7_4

Cheng, S., & Werning, M. (2016). What is episodic memory if it is a natural kind? *Synthese*, *193*(5), 1345–1385. https://doi.org/10.1007/s11229-014-0628-6

Cheng, S., Werning, M., & Suddendorf, T. (2016). Dissociating memory traces and scenario construction in mental time travel. *Neuroscience & Biobehavioral Reviews*, *60*, 82–89. https://doi.org/10.1016/j.neubiorev.2015.11.011

Diba, K., & Buzsáki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience*, *10*(10), 1241–1242. https://doi.org/10.1038/nn1961

Fang, J., Demic, S., & Cheng, S. (2018). The reduction of adult neurogenesis in depression impairs the retrieval of new as well as remote episodic memory. *PLoS One*, *13*(6), e0198406. https://doi.org/10.1371/journal.pone.0198406

Fang, J., Rüther, N., Bellebaum, C., Wiskott, L., & Cheng, S. (2018). The interaction between semantic representation and episodic memory. *Neural Computation*, *30*(2), 293–332. https://doi.org/10.1162/neco_a_01044

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47.

Folkerts, S., Rutishauser, U., & Howard, M. W. (2018). Human episodic memory retrieval is accompanied by a neural contiguity effect. *Journal of Neuroscience*, *38*(17), 4200–4211. https://doi.org/10.1523/JNEUROSCI.2312-17.2018

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

Foster, D. J., & Wilson, M. a. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, *440*(7084), 680–683. https://doi.org/10.1038/nature04587

Franzius, M., Sprekeler, H., & Wiskott, L. (2007). Slowness and sparseness lead to place, head-direction, and spatial-view cells. *PLoS Computational Biology*, *3*(8), e166. https://doi.org/10.1371/journal.pcbi.0030166

Franzius, M., Wilbert, N., & Wiskott, L. (2011). Invariant object recognition and pose estimation with slow feature analysis. *Neural Computation*, *23*(9), 2289–2323. https://doi.org/10.1162/NECO_a_00171

Ganguli, S., & Sompolinsky, H. (2012). Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annual Review of Neuroscience*, *35*(1), 485–508. https://doi.org/10.1146/annurev-neuro-062111-150410

Girardeau, G., Benchenane, K., Wiener, S. I., Buzsáki, G., Zugaro, M. B. M. B., Buzsaki, G., & Zugaro, M. B. M. B. (2009). Selective suppression

of hippocampal ripples impairs spatial memory. *Nature Neuroscience*, *12*(10), 1222–1223. https://doi.org/10.1038/nn.2384

Graham, K. S., Barense, M. D., & Lee, A. C. H. (2010). Going beyond LTM in the MTL: A synthesis of neuropsychological and neuroimaging findings on the role of the medial temporal lobe in memory and perception. *Neuropsychologia*, *48*(4), 831–853. https://doi.org/10.1016/j.neuropsychologia.2010.01.001

Gross, C. G. (1992). Representation of visual stimuli in inferior temporal cortex. *Philosophical Transactions of the Royal Society B*, *335*(1273), 3–10. https://doi.org/10.1098/rstb.1992.0001

Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S., & Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. *Neuron*, *65*(5), 695–705. https://doi.org/10.1016/j.neuron.2010.01.034

Howard, M. W., & Eichenbaum, H. (2013). The hippocampus, time, and memory across scales. *Journal of Experimental Psychology General*, *142*(4), 1211–1230. https://doi.org/10.1037/a0033621

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, *148*(3), 574–591.

Ji, D., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience*, *10*(1), 100–107. https://doi.org/10.1038/nn1825

Kompella, V. R., Luciw, M., & Schmidhuber, J. (2012). Incremental slow feature analysis: Adaptive and episodic learning from high-dimensional input streams. *Neural Computation*, *24*(11), 2994–3024. https://doi.org/10.1162/NECO_a_00344

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language*, *37*(37), 555–583. https://doi.org/10.1006/jmla.1997.2529

Lech, R. K., & Suchan, B. (2014). Involvement of the human medial temporal lobe in a visual discrimination task. *Behavioural Brain Research*, *268*, 22–30. https://doi.org/10.1016/j.bbr.2014.03.030

Lee, A. C. H., Bandelow, S., Schwarzbauer, C., Henson, R. N. A., & Graham, K. S. (2006). Perirhinal cortex activity during visual object discrimination: An event-related fMRI study. *NeuroImage*, *33*(1), 362–373. https://doi.org/10.1016/j.neuroimage.2006.06.021

Lee, A. C. H., Buckley, M. J., Pegman, S. J., Spiers, H., Scahill, V. L., Gaffan, D., … Graham, K. S. (2005). Specialization in the medial temporal lobe for processing of objects and scenes. *Hippocampus*, *15*(6), 782–797. https://doi.org/10.1002/hipo.20101

Lee, A. C. H., Fischer, T. J., Murray, E. A., Saksida, L. M., Epstein, R. A., Kapur, N., … Graham, K. S. (2005). Perceptual deficits in amnesia: Challenging the medial temporal lobe "mnemonic" view. *Neuropsychologia*, *43*(1), 1–11. https://doi.org/10.1016/j.neuropsychologia.2004.07.017

Lee, A. C. H., & Rudebeck, S. R. (2010). Human medial temporal lobe damage can disrupt the perception of single objects. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(19), 6588–6594. https://doi.org/10.1523/JNEUROSCI.0116-10.2010

Lee, A. C. H., Scahill, V. L., & Graham, K. S. (2008). Activating the medial temporal lobe during oddity judgment for faces and scenes. *Cerebral Cortex*, *18*(3), 683–696. https://doi.org/10.1093/cercor/bhm104

Lee, A. K., & Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron*, *36*(6), 1183–1194. https://doi.org/10.1016/s0896-6273(02)01096-6

Lennie, P. (2003). The cost of cortical computation. *Current Biology*, *13*(6), 493–497. https://doi.org/10.1016/S0960-9822(03)00135-0

Levy, W. B. (1996). A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus*, *6*(6), 579–590.

Li, N., & DiCarlo, J. J. (2010). Unsupervised natural visual experience rapidly reshapes size-invariant object representation in inferior temporal cortex. *Neuron*, *67*(6), 1062–1075. https://doi.org/10.1016/j.neuron.2010.08.029

Lin, L.-J. (1993). *Reinforcement learning for robots using neural networks*. (PhD thesis). Carnegie Mellon University. Retrieved from http://isl.anthropomatik.kit.edu/pdf/Lin1993.pdf.

Lisman, J. E. (1999). Relating hippocampal circuitry to function: Recall of memory sequences by reciprocal dentate-CA3 interactions. *Neuron*, *22*(2), 233–242.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577–621. https://doi.org/10.1146/annurev.ne.19.030196.003045

Louie, K., & Wilson, M. A. (2001). Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron*, *29*(1), 145–156. https://doi.org/10.1016/S0896-6273(01)00186-6

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

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., … Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533. https://doi.org/10.1038/nature14236

Pastalkova, E., Itskov, V., Amarasingham, A., & Buzsáki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science*, *321*(5894), 1322–1327. https://doi.org/10.1126/science.1159775

Pyka, M., & Cheng, S. (2014). Pattern association and consolidation emerges from connectivity properties between cortex and hippocampus. *PLoS ONE*, *9*(1), e85016. https://doi.org/10.1371/journal.pone.0085016

Rolls, E. T. (2000). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron*, *27*(2), 205–218. https://doi.org/10.1016/S0896-6273(00)00030-1

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, *2*(9), 437–442. https://doi.org/10.3758/BF03208784

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, *20*(1), 11–21.

Shrager, Y., Gold, J. J., Hopkins, R. O., & Squire, L. R. (2006). Intact visual perception in memory-impaired patients with medial temporal lobe lesions. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(8), 2235–2240. https://doi.org/10.1523/JNEUROSCI.4792-05.2006

Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, *27*, 279–306. https://doi.org/10.1146/annurev.neuro.27.070203.144130

Squire, L. R., & Zola, S. M. (1998). Episodic memory, semantic memory, and amnesia. *Hippocampus*, *8*(3), 205–211.

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*(5026), 1380–1386. https://doi.org/10.1126/science.1896849

Stark, C. E. L., & Squire, L. R. (2000). Intact visual perceptual discrimination in humans in the absence of perirhinal cortex. *Learning & Memory*, *7*(5), 273–278. https://doi.org/10.1101/lm.35000

Suzuki, W. A. (2009). Perception and the medial temporal lobe: Evaluating the current evidence. *Neuron*, *61*(5), 657–666. https://doi.org/10.1016/j.neuron.2009.02.008

Teyler, T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory. *Behavioral Neuroscience*, *100*(2), 147–154.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*(3), 391–412. https://doi.org/10.1088/0954-898X_14_3_302

Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J. J., Moser, M.-B., & Moser, E. I. (2018). Integrating time from experience in the lateral entorhinal cortex. *Nature*, *561*(7721), 57–62. https://doi.org/10.1038/s41586-018-0459-6

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–402). New York, NY: Academic Press, Inc..

Tulving, E. (1974). Cue-dependent forgetting. *American Scientist*, *62*(1), 74–82.

Tulving, E. (1995). Organization of Memory: Quo Vadis? In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (Vol. 839847, pp. 839–847). Cambridge, MA: MIT Press. https://doi.org/10.1017/S0140525X00047257

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, *64*(1), 49–60. https://doi.org/10.1037/h0044616

Wallenstein, G. V., Eichenbaum, H., & Hasselmo, M. E. (1998). The hippocampus as an associator of discontiguous events. *Trends in Neurosciences*, *21*(8), 317–323. https://doi.org/10.1016/S0166-2236(97)01220-4

Wimmer, G. E., & Shohamy, D. (2012). Preference by association: How memory mechanisms in the hippocampus bias decisions. *Science*, *338*(6104), 270–273. https://doi.org/10.1126/science.1223252

Wiskott, L., & Sejnowski, T. J. (2002). Slow feature analysis: Unsupervised learning of invariances. *Neural Computation*, *14*(4), 715–770. https://doi.org/10.1162/089976602317318938

Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*, *75*(1), 168–179. https://doi.org/10.1016/j.neuron.2012.05.010

Zito, T., Wilbert, N., Wiskott, L., & Berkes, P. (2008). Modular toolkit for data processing (MDP): A python data processing framework. *Frontiers in Neuroinformatics*, *2*, 1662–5196. https://doi.org/10.3389/neuro.11.008.2008

How to cite this article: Görler R, Wiskott L, Cheng S. Improving sensory representations using episodic memory. *Hippocampus*. 2019;1–19. https://doi.org/10.1002/hipo.23186