

## **1. Technical Approach**

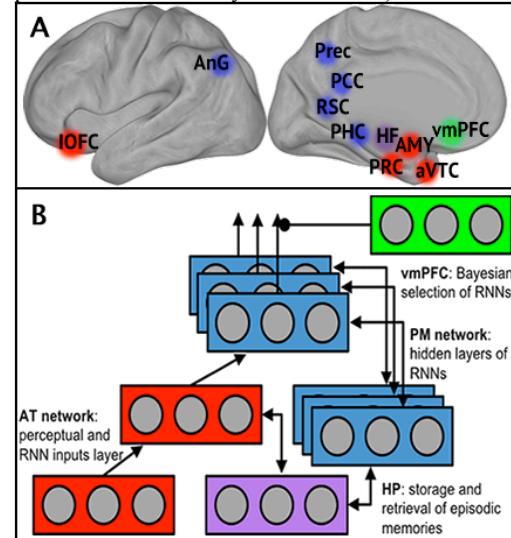
### **1a. Cognitive and Neural Mechanisms of Event Representation and Episodic Memory**

People break the sensory stream into discrete, temporally structured events (“**event segmentation**”). For instance, while watching movies of everyday activities, observers show a striking degree of agreement about the beginning and endpoints of depicted events (i.e., “**event boundaries**”)<sup>1,2</sup>, and people appear to segment events automatically<sup>5</sup>. According to Event Segmentation Theory (EST)<sup>7</sup>, these findings reflect the fact that people maintain a mental model (the “**current event representation**”) that specifies not only the current situation, but also what has happened and what could happen within the boundaries of the event. The current event representation is constructed using semantic knowledge about particular types of events that have been experienced in the past (“**event schemas**”). Recent findings indicate that event structure shapes both memory for the past and anticipation of the future — recall of past information and prediction of future events are highly accurate within the boundaries of an event, but even when the passage of time is equated, recall and prediction are much worse across event boundaries<sup>12</sup>. *These findings indicate that any viable model of memory needs to account for the consequences of event structure.*

Human memory is supported by two cortico-hippocampal networks (Fig. 1). In the 50 years since the initial report of Patient H.M.<sup>13</sup>, neuroscience research on memory has focused on the hippocampus. Since that time, the field has converged on two general points: (1) the hippocampus plays a critical role in associating specific information about people and things with information about the context in which they were encountered<sup>14-19</sup>, and (2) the hippocampus supports learning of event-specific information, but it does not represent remotely acquired semantic knowledge<sup>20</sup>. Human neuroscience research further indicates that the hippocampus primarily interacts with two different neocortical networks: a **Posterior Medial (PM) network** (aka the “default network,”<sup>21</sup> Fig. 2a) that includes the parahippocampal (PHC), posterior cingulate, retrosplenial, and ventromedial prefrontal cortex, along with the precuneus and the angular gyrus; and an **Anterior Temporal (AT) network** that includes the perirhinal, lateral orbitofrontal and ventral temporopolar cortex, and the amygdala<sup>3,22</sup>.

Regions in the PM network contribute to event representation. PM regions consistently co-activate with the hippocampus during recall of real-life events, recollection of the context associated with studied items (Fig. 2b), retrieval of temporal sequences, imagination of future events, and spatial navigation<sup>3,23,24</sup>. These tasks all require one to gradually construct a mental representation of an event. Consistent with this idea, fMRI and electrocorticography (ECoG) studies by Hasson and colleagues<sup>25</sup> have shown that *regions in the PM network integrate information across hundreds of seconds* (Fig. 2c); Zacks

**Fig. 1. (A)** The PM/AT framework<sup>3</sup>. **(B)** Diagram of SEM Architecture with corresponding brain regions. The AT network (red) is hypothesized to process information about entities. This information is fed into event schemas, which are instantiated in the PM network (blue) as recurrent neural networks (RNNs). vmPFC (green) is hypothesized to select the currently relevant event schema/RNN. The hippocampus is hypothesized to support storage and retrieval of event-specific information (i.e., activity patterns in the currently-selected RNN).



and colleagues have shown that regions in the PM network show increased activity during event boundaries<sup>26</sup> (**Fig. 2d**).

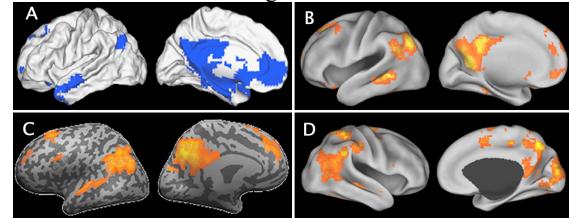
The strongest evidence for PM network involvement in event representation comes from recent work by Baldassano et al.<sup>27</sup>, who used a Hidden Markov Model (HMM) to analyze fMRI data collected as subjects viewed films, heard verbal narratives of those films, and subsequently recalled the films. HMMs are used in neuroscience to identify sequences of neural states that are tied to environmental observations<sup>28,29</sup>; here, we used the HMM to identify sequences of stable neural activity patterns in the PM network that might correspond to neural representations of events. Consistent with EST, the HMM identified a reliable succession of activity patterns in the PM network that was consistent regardless of whether subjects viewed, recalled, or listened to a description of the movie. Moreover, despite the fact that the HMM was only trained on neural data from the PM network, the timing of neural state transitions identified by the HMM corresponded closely to event boundaries identified by independent human observers. *These findings indicate that the PM network is a strong candidate for representing modality-independent event schemas.* Another important result from this study is that activity state transitions in the PM network coincided with abrupt increases in hippocampal activity, and the magnitude of hippocampal activity spikes predicted individual differences in episodic memory for movie-specific information. These results suggest that *any valid neurocomputational model of event representation and episodic memory must be able to account for interactions between the PM network and the hippocampus.*

## 1b. Computational Approaches to Event Representation and Episodic Memory

Interactions between the hippocampus and neocortex mediate a fundamental tradeoff between generalization across events and retention of event-specific information<sup>30</sup>. Simulations within the Complementary Learning Systems (CLS) framework<sup>31</sup> indicate that the architecture of the hippocampal circuit is designed to assign distinct, non-overlapping representations to overlapping input patterns (i.e., “**pattern separation**”), which allows new input patterns to be rapidly learned without causing catastrophic interference between memories. During subsequent encounters with a degraded version of a previously encountered input pattern, the hippocampal model rapidly regenerates the entire input pattern (“**pattern completion**”), thereby supporting recollection of event-specific information. In the CLS model, neocortical areas assign overlapping representations to similar inputs, supporting schema-like<sup>32</sup> learning of statistical regularities across multiple experiences. A key insight from CLS simulations is that *the hippocampus is needed to represent event-specific information, and that this information can be incorporated into neocortical schemas via reactivation of hippocampal representations during online cognition or offline sleep rest states<sup>30,33</sup>.*

Schemas can be modeled as prediction machines. EST is based on the premise that event schemas are adaptive because they generate inferences and predictions about the state of the world. Reynolds et al.<sup>34</sup> computationally implemented this idea using a recurrent neural network (RNN), a model that is optimized for extracting environmental statistics to predict future events. Activity patterns in the hidden layer of the RNN reflect the currently active event representation,

**Fig. 2.** The PM network supports event cognition. Images depict regions showing: (A) High intrinsic functional connectivity with the hippocampus<sup>4</sup>, (B) Increased activity during conscious recollection<sup>6</sup>, (C) Modality-invariant activation during narrative comprehension<sup>8</sup>, (D) Enhanced activation during event boundaries<sup>1,9-11</sup>.



and its connection weights can be viewed as the event schema (i.e., knowledge about what usually happens for this type of event). Information in the input layer of the RNN is integrated with information from the recent past, via activity patterns in the hidden layers. The connection weights, in turn, capture statistical regularities across similar events, such that input patterns drive the hidden layer to generate a prediction about the next sensory input in the output layer. When environmental input deviates from the RNN's prediction, this induces a gating signal that updates the currently active event representation. *A critical strength of this RNN approach is that it is capable of capturing long-range sequential dependencies in an unsupervised fashion.*

Latent cause models explain schema learning and inference. Over a lifetime, people accumulate a massive library of event schemas, and this raises the question of how to select appropriate event schema at the present moment ("schema inference"). Latent Cause Models (LCMs)<sup>35,36</sup> solve this problem by using probabilistic inference either to activate the existing schema with the lowest prediction error, or to create a new schema if overall prediction error is too high. Once a schema is inferred to be currently relevant, it is updated with current input<sup>37</sup>, and, as the schema becomes populated with experiences, it comes to represent semantic knowledge about statistical regularities associated with particular kinds of events. Recent evidence indicates that a noisy neural population code can approximate these probabilistic computations<sup>37</sup>. *The LCM framework provides a simple, biologically plausible computational mechanism for how the brain retrieves or creates the most relevant schema after an event boundary, and for the conditions under which a schema will be modified (reconsolidated) by experience.*

Models of event knowledge need to represent objects, entities, and relations amongst them<sup>38,39</sup>. Although people naturally represent events in terms of relationships (e.g., "Joe ate a hamburger" ≠ "A hamburger ate Joe"), most computational approaches to memory and perception represent events in terms of feature conjunctions that are insensitive to role relationships. Although a deep neural network can be trained to disambiguate narrow, clearly defined relationships<sup>40</sup>, a different approach is needed in order to represent knowledge that can support an infinite number of inferences in different contexts. Early models<sup>41</sup> accounted for relational structure in the form of propositions that specify relationships between entities. Unfortunately, these symbolic models lack neural plausibility, are limited in their ability to tolerate deviations in event content (e.g., "Joe ate a cheeseburger"), and lack a mechanism for flexible learning and updating of event schemas based on specific experiences. We have addressed the limitations of structured symbolic representations by operationalizing them within a distributed neural representation<sup>35,42-46</sup>. *The structured representation approach provides a powerful method for representing events in a neurally plausible manner that can support human-like inferences and predictions.*

### 1c. Research Challenges

Available evidence indicates: (1) People segment experiences into meaningful, structured events, and event segmentation has significant consequences for memory and online processing. *We lack an adequate understanding of the factors that drive event segmentation, and we know little about whether or how memory consolidation and retrieval changes event representation.* (2) The hippocampus is ideally suited to store memories for event-specific details, whereas the PM network is likely to store knowledge about event schemas that generalize across classes of specific events. *We know little, however, about how cortico-hippocampal interactions relate to event segmentation, schema inference, and memory consolidation.* (3) CLS provides a detailed computational account of hippocampal function, RNNs can capture event schema representation, and LCMs provide a computational account for schema inference. *There is no precedent,*

however, for a coherent model of episodic memory and event representation that captures the relational structure of real-life events. As we describe below, our separate lines of research have laid the groundwork for an ambitious, integrated computational cognitive neuroscience research program that will fundamentally advance our understanding of event representation and episodic memory.

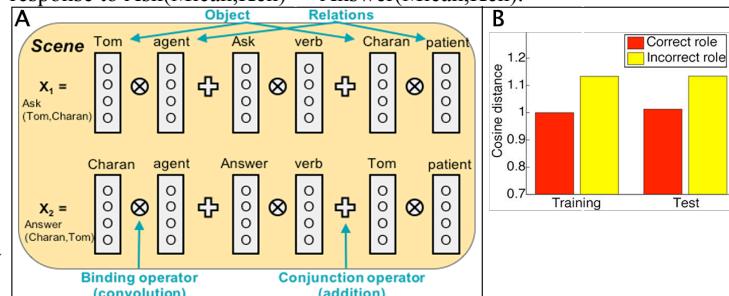
#### 1d. The Structured Event Memory (SEM) Computational Framework

SEM leverages the strengths of the CLS, LCM, and EST models in a coherent architecture that is directly linked to the functional organization of cortico-hippocampal networks outlined in the PM/AT framework (**Fig. 1b**). At the core of SEM, modern large-scale RNNs are employed to model event schemas that capture the rich structure of real-world events. This library of RNNs is linked to the CLS hippocampal model, which enables the RNNs to reconstruct activity patterns elicited by specific past experiences. Mechanisms from LCMs are used to regulate the creation, selection, and updating of RNNs during schema inference and event segmentation. Critically, SEM is designed to handle structured representations that preserve relationships between entities. As described below, SEM provides a parsimonious platform for generating specific testable hypotheses (labeled in **BLUE** throughout) about a wide range of phenomena, including (but not limited to) event segmentation, construction of event models, the distinction between episodic and semantic memory, and memory consolidation and reconsolidation.

Given that SEM is inspired by, and designed to account for, neuroscience data, we have formulated predictions about the correspondence between the model and activity in different brain areas. Our initial assumption is that the AT network encodes semantic information about objects that is instantiated in the input layers of the RNNs in SEM. Based on evidence implicating the PM network in representing features of events<sup>3,25,27</sup>, we can generate predictions about PM activity based on activity in hidden layers of the RNNs in SEM. We propose that ventromedial prefrontal cortex (vmPFC) encodes the posterior probability distribution over event schemas (i.e., which schemas are most probable given the observed data), based on our recent findings<sup>47</sup>, along with other work implicating vmPFC in the representation of the current state<sup>48,49</sup>. Finally, we predict that the *midbrain dopamine system* (ventral tegmental area and substantia nigra)<sup>50,51</sup> and *locus coeruleus*<sup>52</sup> signal prediction errors and modulate hippocampal encoding<sup>53,54</sup>. Although all of these predictions are strongly motivated by the existing literature, our empirical data will allow us to discover alternative anatomical linking assumptions.

**Structured representation in SEM:** We propose that SEM receives processed input about the current state of the world in the form of a “**scene**.” Each scene vector is composed of vector representations for constituent objects, relations between objects, a binding operator (which grounds relations by composing object and relation vectors) and a conjunction operator (which computes the vector representation for a scene by composing grounded vectors).

**Fig. 3. (A)** Holographic reduced representations in SEM. **(B)** Left: SEM is trained and later tested (red) with a sequence of scenes: Ask(Tom,Charan) -> Answer(Charan, Tom). SEM generates a large prediction error (i.e., increased cosine distance) if the correct role-filler binding is violated (yellow) by presenting Answer(Tom, Charan). Right: SEM generates a relatively small prediction error (red) when tested on the sequence Ask(Micah,Ken) -> Answer(Ken,Micah), showing that it can generalize across role fillers. However, it generates a large prediction error in response to Ask(Micah,Ken) -> Answer(Micah,Ken).



These operations can be implemented in several ways<sup>55-57</sup>; our current prototype (see **Task 1**) employs holographic reduced representations that use circular convolution as the binding operator and vector addition as the conjunction operator. This produces vectors of fixed dimensionality regardless of the complexity of the underlying scene. The underlying components can be approximately decoded using circular correlation<sup>56</sup> (**Fig. 3**). Simulations summarized in **Fig. 3b** show that *this formalism supports event representations that are structure-sensitive and generalize across variables*. One important property of holographic reduced representations is that the resulting vectors are differentiable, allowing efficient learning algorithms based on gradient descent. We do not claim that the brain literally computes circular convolution or correlation, but possibly some approximation of these operations<sup>58</sup>.

Semantic memory in SEM: SEM distinguishes between semantic knowledge about relations *within* events, and semantic knowledge about relations *between* events. Semantic knowledge about relationships between scenes in an event is represented via connection weights in individual RNNs (e.g., during a dishwashing event, “apply soap” is usually followed by “rinse”). Knowledge about relationships between different classes of events (e.g., “dinner” events tend to be followed by “dishwashing” events) is represented within the event transition probability distribution in the LCM. In short, semantic event knowledge in SEM can be viewed as a library of RNNs that accumulate detailed information about recurring event schemas, along with machinery for inferring which RNN is currently relevant.

Prediction errors in SEM: The difference between the RNN’s output layer pattern and the input pattern at the next time step is called the **prediction error (PE)**. At every time step, the PE is used to update the currently active RNN by using a learning algorithm (like gradient descent) that alters model weights in order to reduce prediction errors over time. Over the course of training, the RNN learns to maintain information in its hidden layers that is useful for predicting the properties of upcoming scenes. The hidden layers carry useful information from the current input scene pattern and diagnostic information from other recently processed scenes (e.g., a character who is present in the room but not currently visible).

Large prediction errors drive event segmentation (H1). For instance, if SEM runs an RNN for “Meeting” events during a teleconference between Tom and Charan, the meeting RNN should be able to predict an upcoming scene like “Tom asked Charan a question” [i.e., Ask(Tom, Charan)]. However, the RNN’s predicted output would be wrong if the next scene is Drive(Tom, Supermarket); if this happens, the LCM will likely infer an event boundary between the Ask scene and the Drive scene and either select a different event schema RNN (e.g., Shopping Trip) or create a completely new RNN (see “Schema Inference” below). Because prediction errors drive event segmentation in SEM, the model predicts that *one should see fewer event boundaries in entirely unfamiliar or unpredictable situations than in situations that are well-captured by existing schemas (H2)*. This is because a newly initialized RNN or an RNN trained on unpredictable stimuli will generate weak predictions that are likely to correspond to many possible scenes, leading to weak prediction errors. In contrast, with extensive training, an RNN will generate strong predictions about what comes next, so it will generate large prediction errors when these strong predictions are violated, leading to event segmentation. Also, *sudden changes in input patterns over time will be more likely to trigger event segmentation than gradual changes, because the former are more likely to trigger a large prediction error (H3)*<sup>36,59</sup>.

Schema inference. The LCM component of the model generates predictions about upcoming events (the “prior over events”) using the “sticky Chinese restaurant process” (sCRP). The sCRP

is a simple Bayesian nonparametric prior that operationalizes the assumptions that (1) higher frequency events are more likely to be repeated, (2) the most recent event is more likely to remain active, thereby inducing continuity, and (3) there is a non-zero probability of generating new events. Formally, the sCRP is calculated as follows: at time  $n$  the next event is drawn from  $P(e_n = k | \mathbf{e}_{1:n-1}) \propto C_k + \lambda \mathbb{I}[e_{n-1} = k]$  if event  $k$  was previously active.  $C_k$  is the number of previous timepoints assigned to event  $k$ , and  $\mathbb{I}[\cdot] = 1$  if its argument is true (0 otherwise). With probability proportional to  $\alpha$ , a new event is created (i.e., one that was not previously active). The stickiness parameter  $\lambda$  determines the degree of temporal continuity of events.  $\lambda$  embodies the assumption in SEM that meaningful events are continuous in time, such that – at typical values – the model will tend to adopt priors that favor maintenance of the same event schema. The concentration parameter  $\alpha$  determines the simplicity bias; at typical values,  $\alpha$  favors fewer distinct events. To account for event schemas, we will use values of the stickiness and concentration parameters that optimize representation of events at a coarse timescale, and generalization across multiple events.

Information incorporated in the prior is integrated with new information about the *likelihood* of each event schema, which is computed by taking the inverse of the prediction error generated by each RNN (i.e., higher PE = lower likelihood that the event “generated” the current input). Bayes’ rule is used to integrate the prior with the likelihood to identify the event schema with the highest posterior probability. Neurally, this corresponds to the selection of a particular cortical RNN for training using the current input. If no existing RNN is sufficiently likely, then the model generates a new RNN that is trained with the current input. Thus, SEM learns from every new input pattern, either by updating existing schemas or creating a new one.

Because prediction errors drive both schema inference and learning in the inferred RNN, the model predicts that *event segmentation determines schema updating* (**H4**): If a PE triggers event segmentation, the newly selected RNN will be trained, and if segmentation does not occur, the currently active RNN is trained. SEM also predicts that *segmentation outcomes will be driven by features that are highly diagnostic of event identity* (**H5**). For example, if I always eat Pancakes at Breakfast, then the appearance of Pancakes will lead the LCM to select the Breakfast RNN, and its disappearance will lead the LCM to infer an event boundary.

Modeling episodic memory via cortico-hippocampal interactions: Contrary to many popular models in cognitive neuroscience<sup>60,61</sup>, we do not equate episodic or declarative memory with the hippocampus. Instead, SEM builds on work showing that episodic memory is heavily schema-driven<sup>32</sup>, and it formalizes the assumption in the PM/AT framework<sup>22,62</sup> that interactions between the hippocampus and PM network support episodic encoding and retrieval. In SEM, memory for specific events is supported by interactions between the RNNs and the biologically-realistic hippocampal neural network architecture implemented in the CLS model<sup>31,63</sup>.

Prediction errors regulate episodic encoding in SEM. Motivated by evidence suggesting that prediction errors (or “novelty”) facilitate hippocampal memory formation and consolidation<sup>51,54,64</sup>, and by our findings showing that hippocampal activity magnitude<sup>27,65</sup> and activity pattern change<sup>66</sup> is increased during event boundaries, we predict that *the hippocampus encodes event-specific information (“snapshots”) at moments of high prediction error in PM-network structures* (**H6**). During a snapshot, the hippocampal model encodes a representation of the currently active RNN, along with a record of its current contents, including the input pattern (i.e., the scene vector) and the pattern of hidden layer activation. In other words, hippocampal snapshots carry all of the necessary information for reinstating a past event, including the event

context and the temporal, causal, and situational relationships between the entities that were encountered.

Cortico-hippocampal interactions regulate episodic retrieval in SEM. Because the model is designed to reconstruct input patterns when given a partial cue, SEM predicts that *hippocampal retrieval of RNN states will be triggered when the RNN's activity patterns are a partial match with previous experiences* (i.e., there are some matching features and no strongly mismatching features) (**H7**). Partial match states will be especially likely to occur at moments when the RNN processes a previously encountered input pattern but cannot generate a strong prediction of the next input (i.e., high **entropy**) – this inability to predict the next state effectively causes a gap in the cortical pattern that the hippocampus can fill in. Because retrieved memories are cycled through the RNN, SEM does not veridically reproduce past experiences; rather, *event representations are biased toward reconstructions that are schema-consistent* (**H8**)<sup>32</sup>.

Reactivation updates event schemas. Because hippocampal output serves as a source of input to the RNN, hippocampal pattern completion can trigger learning in the RNN. This leads to a number of interesting predictions: In particular, *if an input drives pattern completion, its association with related event-specific information should be enhanced* (**H9**). For example, if scene A occurred before scene B previously, then if A triggers pattern completion of B, this should strengthen the RNN's ability to predict B. As a corollary, *hippocampal retrieval can drive updating or “reconsolidation” if new information is incorporated into an RNN shortly after pattern completion*<sup>67</sup> (**H10**). Finally, in the SEM account, reactivation of recently learned information during offline states such as sleep or waking rest works like a stochastic gradient descent algorithm – by randomly reactivating recently acquired hippocampal representations, the network trains itself to reduce prediction error. In other words, *sleep or rest should facilitate the ability to associate linked information within an event schema* (**H11**).

## 1e. Overview and General Methods

**Table 1** summarizes the numbered hypotheses that will be tested in the empirical research studies (Tasks 2-5). Methods common across studies are summarized below.

**Table 1.** Hypotheses that will be tested in Tasks 2-5.

Task 2	Task 3	Task 4	Task 5
<b>H2-8</b>	<b>H7, 9-11</b>	<b>H1, 6-7</b>	<b>H1, 6-7</b>

fMRI acquisition, processing, and analysis: Scanning protocols will be continually optimized across sites, but in all studies we will use a multiband gradient-echo echo planar imaging (EPI) pulse sequence (<https://www.cmrr.umn.edu/multiband/>)<sup>68</sup> with whole-brain coverage, at least 3.0mm isotropic spatial resolution and at least 1.2s temporal resolution. Echo time, slice orientation, and orientation/size of the shim volume are chosen to minimize susceptibility artifacts and maximize BOLD contrast in the hippocampus, AT network, and vmPFC. Field maps, T1-weighted MPRAGE images (1mm isotropic resolution), and contrast-matched segmented EPI images are also collected. Total duration of an entire scan session is typically 90-120 minutes. Reconstructed EPI images are motion-corrected, coregistered with corresponding structural scans, and normalization parameters are calculated using the DARTEL procedure in SPM12. Univariate General Linear Model (GLM) analyses will be implemented in either SPM12 or FSL packages. Motion spikes will be detected with the Artifact Detection Toolbox ([http://www.nitrc.org/projects/artifact\\_detect/](http://www.nitrc.org/projects/artifact_detect/)) and scrubbed<sup>69,70</sup> by including nuisance regressors in the GLM. Left- and right- hippocampal regions of interest (ROIs) will be manually traced on high-resolution structural images, and ROIs for PM and AT network regions will be delineated from whole-brain maps of intrinsic functional connectivity with the posterior and anterior

hippocampus based on large independent datasets from the Ranganath lab<sup>4,71,72</sup>. We can also interrogate structural ROIs derived from the Destrieux et al. atlas<sup>73</sup> (see <sup>6,74</sup> for examples).

**Multivariate fMRI analysis methods:** We will use *Representational Similarity Analysis* (RSA)<sup>75</sup> to test whether a brain region carries information about a particular kind of stimulus or memory attribute by contrasting similarity in neural patterns associated with pairs of trials that share the same attribute value against pairs of trials that involved dissimilar attributes. RSA will be performed by first estimating single-trial activity for each voxel<sup>76,77</sup> (multi-trial modeling is also an option and produces comparable results), and, for each trial, the set of Beta values is concatenated into a “voxel pattern” vector. Pearson’s  $r$  is used to quantify pattern similarity (PS) between pairs of trials and the resulting correlation coefficients are selectively averaged, Fisher-transformed, and entered into group multilevel mixed effects models that also incorporate regressors to control for nuisance effects such as the passage of time. We will also use *Hidden Markov Model (HMM)* analyses to test predictions about neural representations of event structure in a data-driven manner. Specifically, our HMM analyses model a neural time series by assuming that it was generated by a sequence of discrete transitions between stable, event-specific neural activity patterns (fMRI in Tasks 2-4; ECoG in Task 5). Fitting the HMM entails estimation of: (1) the timing of transitions between events, (2) the activity pattern unique to each event representation, and (3) the best-fitting number of events. For more details on how the HMM is fit, see Baldassano et al (2016)<sup>27</sup>. *After the HMM is fit to a neural time series, it can be used to identify corresponding neural event representations in a new dataset, even if the length of the new dataset and/or the timing of underlying events within the dataset differs from the training data.* Finally, as described below in **Task 1**, we will use SEM to generate a theory-driven forward model to decode cognitive states from fMRI and ECoG data.

### **Task 1: Develop a Computational Model of Event Representation and Episodic Memory (S. Gershman, Harvard)**

**Subtask 1: Design and Evaluate the Neural Network Architecture of SEM.** We have implemented a prototype version of SEM (available at <https://github.com/sjgershm/SEM>), using a linear network with no hidden layer. However, we believe that a short-term memory representation and nonlinear dynamics are crucial elements of a successful model of event schemas. We will explore several types of RNNs that have recently achieved state-of-the-art performance in artificial intelligence systems, such as long short-term memory networks and differentiable neural computers<sup>78,79</sup>. We will evaluate RNN architectures in three ways. First, we will run the models through simulations that stress-test important cognitive functions, such as the ability to maintain or reinstate information in memory, and the ability to encode structure-sensitive transitions between vector representations. Second, we will identify which RNNs do the best job of emulating human performance on our experimental tasks. Third, we will compare them in terms of biological plausibility.

A second central component of SEM is the compositional vector representation. We are currently using holographic reduced representations (**Fig. 3**), but we will evaluate other possibilities, such as parameterized composition and conjunction operators (an approach that has been successful in recent AI research<sup>80</sup>). We will evaluate different choices of composition architectures in the same way that we evaluate different RNN architectures, as described above.

**Subtask 2: Develop and Validate Forward Models of Neural Data.** We will develop a set of forward models to translate between neural data and cognitive representations in SEM (RNN

activation state and posterior over event schemas). For any choice of mapping, Bayes' rule will be used to invert the forward model and decode cognitive states from neural data. We will start with the simplest forward model, which is equivalent to learning a regularized linear mapping between neural data (multivoxel patterns for fMRI, multi-electrode spectrotemporal patterns for ECoG) and cognitive states identified by SEM. We will also explore alternative models, including the Shared Response Model (SRM), a newly developed technique that maps participants' data into a shared neural space before the decoder is applied. We have found that SRM massively improves fMRI decoding performance for narrative stimuli, by allowing us to productively leverage other participants' data to train the classifier<sup>81,82</sup>; the same technique can be applied to ECoG. To validate these forward models on the data collected at other sites, we will use standard cross-validated predictive metrics (training and testing the forward models on separate datasets) and Bayesian model comparison techniques.

**Subtask 3: Neurally Plausible Implementation of SEM.** We will develop an implementation of SEM that draws upon known anatomy, physiology, and biophysics of the neural circuits implicated in event cognition. Modeling of the hippocampus will be based on the CLS framework<sup>31</sup>; our novel contribution will be to integrate this model within the SEM architecture. The cortical RNNs will be based on known neocortical circuit properties and the physiology of PM networks identified in Tasks 2-5. Finally, we will develop a model of the vmPFC that uses connectivity with the PM network to adaptively gate cortical event representations based on their relevance. We hypothesize that intrinsic network properties (Poisson spiking neurons organized into stable attractors, divisively normalized by inhibitory interneurons) will be sufficient to implement a gating system that approximates the Bayesian computations posited by SEM.

**Subtask 4: Link SEM to Visual and Linguistic Inputs.** In order to scale SEM to real-world sensory inputs, we will develop two elaborated versions of the architecture. First, we will train deep neural networks to learn the mapping from visual (pixel-level) inputs to scene vectors. This will allow SEM to operate on natural images and videos. Second, we will train deep neural networks to learn the mapping from word sequence inputs to scene vectors, allowing SEM to operate on natural language inputs like stories and other narrative media. We will demonstrate the effectiveness of these elaborated architectures by applying them to large-scale datasets of movies and stories, using predictive metrics (e.g., how well can the next image or sentence be predicted) for quantitative assessment and model comparison.

## **Task 2: Specify the Neural Mechanisms That Support Learning and Application of Semantic Knowledge about Events** (K. Norman, U. Hasson, Princeton)

This task will test SEM's predictions about *schematic knowledge*: how it is learned (Exp. 1), how it is represented in the brain (Exp. 2), how it is updated in response to different kinds of changes in the world (Exps. 3 & 4), and how it shapes episodic memory (Exps. 5 & 6).

**Experiment 1: Initial learning of schemas.** SEM predicts that the temporal structure of events is critical for schema learning. SEM learns to differentiate schemas faster if it is given extensive experience with one schema before switching to the next. This training procedure allows the model to build up strong predictions about what will happen within the schema; after this learning has occurred, the model generates large prediction errors when a new schema is shown, leading to (correct) inference of an event boundary (**H2**). To test this, we will contrast neural activity associated with schemas that have been learned in a blocked order (allowing them to build up clear expectations) or an interleaved order. **Stimuli:** We will use a modified version of

the computer program *Sims 3* to generate a series of computer-generated videos depicting a hypothetical "Coffee Shop World"<sup>83</sup>. The videos depict certain kinds of events (e.g., poetry recital; book club meeting; birthday party; fight; coffee tasting), and the succession of scenes within an event type are governed by probabilistic rules; for example, if Attacks(X, Y) then there might be a 60% chance of Attacks(Y, X) and a 40% chance of RunsAwayFrom(Y, X). The *schema* for a particular event consists of all of the rules that govern the succession of scenes for that event. **Procedure:** Participants will incidentally learn the schemas by watching a series of videos during an unscanned training phase (Phase 1). They will not be told the rules or when event boundaries occur. In the *blocked* condition, during Phase 1, participants will observe multiple instances of one type of event (e.g., poetry readings) at the coffee shop, each time with different actors playing different roles; next, participants will observe another type of event (e.g., coffee tastings), and so on. After 3-4 event types in this blocked fashion, we will cycle through all of the event types multiple times so participants get used to frequent switches between event types. Then in Phase 2, we will scan participants as they view different events (with frequent switches between event types) depicted in a continuous video. The *interleaved* condition will be identical except that events from different schemas will be interleaved instead of blocked during Phase 1. We will use the HMM<sup>27</sup> to identify event boundaries in PM network regions while subjects watch the videos, and we will use activity in the Substantia Nigra/VTA as a neural index of prediction error. **Predictions:** The key prediction is that event boundaries identified by the HMM and "spikes" in neural measures of prediction error are expected to coincide more closely with the actual switch between event types in the blocked condition than in the interleaved condition.

**Experiment 2: Neural representation of schematic knowledge.** **Procedure** will be the same as the blocked condition from Exp. 1, except each narrative will be shown in two ways: via a computer animated video and via a verbal narrative (prose). **Predictions:** SEM predicts that the PM network represents event identity (e.g., a poetry reading) and other aspects of the current event representation (e.g., the identity of the poet) in a modality-independent manner. We can test this modality-independence by determining whether neural event representations identified by an HMM generalize across video and spoken media<sup>27</sup>. By contrast, we predict that AT-network regions will contain *modality-dependent* representations of current perceptual inputs (e.g., the neural pattern for computer-animated Jeff will not generalize to the verbal stimulus "Jeff"). We should also be able to decode *predictions about upcoming sensory inputs* from AT-network regions, using the forward model from Task 1. For example, if – at poetry readings – the emcee's introduction is always followed by the featured poet, and we know that the featured poet is Jeff, then it should be possible to decode pre-activation of the Jeff representation in AT-network regions after the emcee appears, but before Jeff actually shows up. Lastly, we will use SEM to compute posterior probabilities over event schemas, and then use RSA to test the prediction that vmPFC should represent the posterior probability distribution over event types. As in Chan et al.<sup>47</sup>, we will manipulate the statistical structure of the stimuli (e.g., varying the degree to which the presence of an emcee will distinguish between poetry readings and coffee tastings), thereby ensuring sufficient variability in the posterior probability distributions.

**Experiment 3: Effects of stimulus diagnosticity on schema updating.** We will test SEM's predictions that 1) the diagnosticity of stimulus features affects event segmentation (**H5**) and 2) event segmentation shapes learning (**H4**). **Procedure:** This study will use a 4-phase design in which we will vary the degree to which an actor or a location (e.g., the decor of the coffee shop) depicted in a video is diagnostic of the corresponding event schema. For example, the presence

of a singer might be diagnostic of birthday party events (i.e., the singer always and only appears at birthday parties). In Phase 1, participants will view several videos outside of the scanner, thereby allowing them to learn the relevant statistics (e.g., at birthday parties, a singer appears after cake is eaten). Next, we will bring participants to the scanner for Phases 2-4. In Phase 2, participants will be scanned as they view videos that are consistent with the previously learned statistics. In Phase 3, we will show participants new examples of videos depicting versions of the previously learned event types (e.g., birthday party) in which event structure is changed by altering either non-diagnostic or diagnostic features (e.g., cake is followed by a clown, not a singer). Finally, in Phase 4, we will show the same videos that we showed during Phase 2.

**Predictions:** We predict that changes to high-diagnosticity people or places during Phase 2 will trigger event boundaries (as measured by the HMM, applied to PM-network regions), thereby leading to updating of the post-boundary schema instead of the pre-boundary schema. We will measure schema updating by comparing predictions elicited during Phase 2 (pre-change) vs. Phase 4 (post-change). In the above example, during Phase 2, participants will expect a singer to appear after the cake; to measure schema updating, we will look at whether participants expect a singer or a clown in Phase 4. We will measure predictions neurally by decoding activation of the singer vs. clown representations in AT-network regions, using the forward model described in Task 1. We will also use behavioral methods for measuring prediction, described in Task 4.

**Experiment 4: Effects of gradual vs. abrupt change on schema updating.** We will test the prediction that gradual change is less likely to trigger event segmentation (and thus more likely to trigger updating) than abrupt change (**H3**). **Procedure:** We will use the same 4-phase design as Experiment 3, except here we will vary (during Phase 3) whether the statistics of an event are changed abruptly or gradually. For example, subjects will learn that, at birthday parties, cake is followed by a singer 25% of the time and by a clown 75% of the time, we can either abruptly flip these statistics to 75% clown, 25% singer or gradually “morph” the probabilities to their new values (progressing through a period where it is 50% singer and 50% clown). **Prediction:** SEM predicts that the gradual change procedure will be less likely to trigger an event boundary during Phase 3 (measured using the HMM) and thus more likely to trigger updating of the changed schema (measured by looking at how predictions change from Phase 2 to Phase 4).

**Experiment 5: Testing how schematic knowledge affects episodic memory storage & retrieval.** We will test SEM’s prediction that the hippocampus takes snapshots of the current event representation when strong prediction errors occur (e.g., at event boundaries; **H6**) and retrieves these snapshots at moments of high uncertainty (e.g., at the start of events; **H7**). **Procedure:** Before scanning, participants will be trained on a set of event schemas by watching videos. One day later, they will be scanned while viewing videos depicting new instances of each learned schema (e.g., a particular birthday party, a particular coffee tasting, etc.). To minimize interference, we will minimize overlap between details of these events (with respect to each other and to videos from the pre-scan training phase) by depicting each event at different spatial contexts (e.g., party in shop A, coffee tasting in shop B, etc. [where A, B, C, and D were not shown during the original unscanned training phase]). Next, participants will be scanned while they re-watch the four events (party in shop A, coffee tasting in shop B, etc). **Predictions:** SEM predicts that hippocampus will take snapshots of events at moments with high prediction error, including (but not limited to) event boundaries (e.g., transitioning from the coffee tasting to the poetry reading). We will use SEM to make theory-driven predictions about the timing of prediction errors, and we will use the HMM to identify event boundaries in PM-network structures in a data-driven manner. We expect that there will be an uptick in hippocampal

activity at high-prediction-error moments (including event boundaries), where the size of the hippocampal activity bump predicts subsequent retrieval when the movies are seen again. We will quantify retrieval during re-watching by using the forward model from Task 1 to decode activity in PM-network regions during re-watching of the events. SEM predicts that retrieval will occur at high-uncertainty moments, which tend to occur at the beginning of the event. If we can decode event features (e.g., identity of barista) before they appear, this effect must be attributable to memory retrieval. Finally, we predict that functional connectivity between the hippocampus and PM-network will be increased during re-watched events, and that the magnitude of this increase will predict decoding success<sup>81</sup>.

**Experiment 6: Testing SEM's predictions about schematic distortion.** This experiment tests SEM's prediction (H8) that episodic memory retrieval will be biased toward reconstructions that are schema consistent (i.e., because retrieved information is cycled through the RNN; H7).

**Procedure:** Same as in Experiment 5, except some of the videos studied in the scanner will violate the pre-trained event schema (e.g., the barista has been Fred in the past, and this time the barista is Joe), and instead of watching each video twice, they will watch only once and subsequently recall the corresponding events in the scanner (as in <sup>84</sup>). **Predictions:** SEM predicts that schematic distortion occurs when the cortical RNN fills in the representation of the default value (Fred); this can happen at encoding (overwriting the representation of Joe) or at retrieval (blocking Joe from coming to mind). To test this prediction, we will use the forward model to track the activation of the default and non-default role fillers during encoding and retrieval, and we will assess when and if the participant "slips" into representing the default value. These neural indicators of schematic distortion should predict behavioral errors during free recall (i.e., reporting that Fred was the barista instead of Joe).

### **Task 3: Specify How Cortico-Hippocampal Interactions Support Episodic Memory Retrieval and Consolidation** (C. Ranganath, UC Davis)

SEM is unique amongst biologically-inspired models of memory, in that it models *interactions* between the hippocampus and the PM network. Here, we test novel predictions about the causes and consequences of cortico-hippocampal interactions for episodic memory.

**Experiment 1: Testing the behavioral prediction that cue-driven retrieval of event-specific information will enhance retention of linked representations within that event. (H9).** In this paradigm, we selectively reactivate memories for scenes from a real-life event, and we examine the effects on later retention of non-reactivated information from the same event. **Procedure:** During Phase 1, each participant will be fitted with a wearable camera that automatically takes photos while s/he attends an educational lecture and demonstration (with live birds of prey) at the California Raptor Center. Each lecture is scripted to consist of several discrete events (verified by independent pilot testing with a videotaped lecture), each including 8 facts about a different bird of prey. Event boundaries for these real-life events were validated in a separate norming study in which participants viewed a video of the lecture and identified event boundaries in the video (e.g., switching from learning about talons to owl pellets). Four hours later, the participants return to the lab to see their photos from the tour (Phase 2). Only some of the Phase 1 photos, selected in order to sample particular moments from each learning event, are shown during Phase 2 (Fig. 4). Participants in Phase 2 will be divided into 2 groups: (1) the Reactivation Group, who will be asked to recall the fact that was presented at the time the photo was taken; (2) the Re-exposure Group, who will be asked to simply rate the quality of the photo.

One week later, the participants will return for **Phase 3** in which they will see photos from Phase 1 that correspond to the scenes where the facts were learned, and they will be asked to verbally recall these facts.

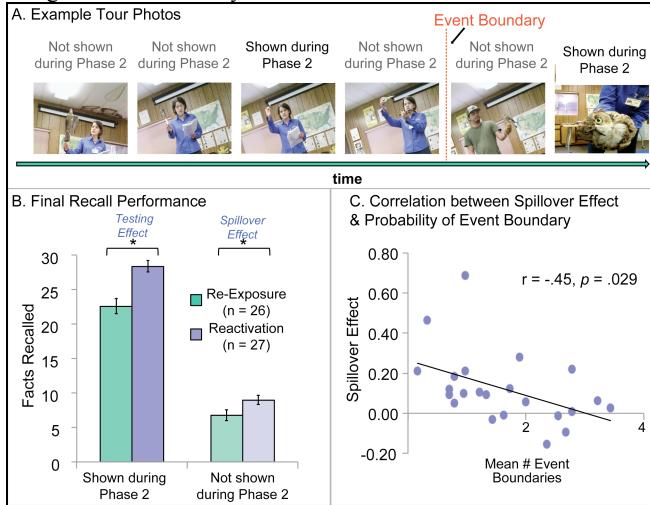
**Predictions:** We expect people to show better memory for facts that were reactivated during Phase 2 than for other facts from the tour. Further, SEM makes the novel prediction that participants in the reactivation group will show better memory for non-reactivated facts than participants in the re-exposure group. This is because recall of scenes is expected to trigger hippocampal pattern completion, thereby reactivating and strengthening other linked scenes from the same event (**H9**). **Preliminary Data** from 54 participants (Fig. 4), revealed that the reactivation group showed better retention of facts corresponding to photos shown in

Phase 2 ( $t(51) = 3.07$ ,  $SE = 0.99$ ,  $p = .003$ ). Critically, the effects of reactivation during Phase 2 “spilled over” to non-reactivated facts from the event ( $t(51) = 2.15$ ,  $SE = 1.02$ ,  $p = .036$ ), consistent with **H9**. Moreover, a detailed analysis revealed that the spillover effect was reduced for scenes that were likely to be separated from the reactivated scene by an event boundary. The preliminary results from Experiment 1 demonstrate that theories of memory consolidation must account for event structure.

**Experiment 2: Testing the prediction that a partial cue will drive reactivation of event representations in the PM network (**H7**), thereby enhancing memory for linked scene information (**H9**).** **Procedure:** This experiment will be similar to Expt. 1 except that fMRI data will be collected during Phase 2 and Phase 3. **Predictions:** SEM predicts that presentation of photos during Phase 2 will drive hippocampal pattern completion, which, in turn, should drive brief reactivation of the neural representation of the corresponding event in the PM network. During Phase 3, presentation of tested and nontested photos should also drive retrieval of this event representation. We will test for reactivation of event representations during Phase 2 by assessing voxel pattern similarity (PS) in PM network regions between Phase 2 reactivation/re-exposure trials and Phase 3 trials. We expect that, in the PM network, PS between scenes that were reactivated during Phase 2 and linked Same-Event scenes from Phase 3 will be higher than PS between reactivated scenes from Phase 2 and Different-Event scenes from Phase 3. We have used RSA to detect spontaneous reactivation of context-specific activity patterns during memory retrieval in many previous studies<sup>6,66,74,85,86</sup>, so we are certain that this study will be feasible. If we find increased PS between reactivated and spillover scenes from the same event, we will run control analyses to rule out effects of perceptual similarity and temporal proximity.

**Experiment 3: Testing whether representations of recent events in the PM network are reactivated during waking rest, and whether offline reactivation strengthens memory for reactivated events (**H11**).** **Procedure:** Participants will be scanned as they watch a video version of the Raptor Center tour. Scanning will continue as they perform a set of Reactivation trials, in

**Fig. 4.** (A) Tour photos from wearable camera. Labels indicate whether an item was reactivated/re-exposed during Phase 2. The red line indicates an event boundary between scenes. (B) Reactivation during Phase 2 improved later retention of facts that were reactivated during Phase 2 (“testing effect”) and also facts that were not cued during Phase 2. (C) The degree to which non-cued items benefited from reactivation was predicted by the likelihood of there being an event boundary between the scenes.



which scenes from the video are used to cue recall of corresponding facts. Each scene will correspond to a unique event in the video, and only half of events will be reactivated during the scanning session. Next, they will complete a 20-minute resting-state fMRI session followed by a scanning session as they watch a video of a different lecture. One week later, participants will complete a final test of memory for all of the facts in the video (i.e., re-activated facts, linked facts from the re-activated event, and facts from non-reactivated events). To test whether neural representations of studied events are reactivated during post-learning rest, we will fit a HMM<sup>27</sup> to fMRI data acquired during the Raptor Tour Video, to identify neural representations of events in the PM network . As in our recent work<sup>71</sup>, we can assess reactivation during rest by using RSA to identify resting-state scans that are similar to neural event representations identified by the HMM during the video. To control for nonspecific effects, we can fit the HMM to fMRI data collected during the second post-rest video and use RSA to compare the event-specific activity patterns with resting state activity patterns<sup>71,87</sup>. **Predictions:** We predict that, within PM network regions, testing of the HMM will reveal significantly more frequent reactivation of Raptor Tour events from the pre-rest video, as compared with events from the post-rest control video. Moreover, we predict that the number of times a particular event is reactivated will predict participants' retention for facts that were associated with the event in the video (**H11**). Finally, an ancillary, but important question is whether the hippocampus is biased to reactivate recently retrieved events. If this is the case, we would expect to see more frequent offline reactivation of events that were cued during the initial memory retrieval/reactivation session than for events that were not reactivated.

**Experiment 4: Testing the prediction that hippocampal retrieval can drive episodic memory updating (“reconsolidation”) if new information is processed shortly after pattern completion (**H10**)**. **Procedure:** As in Expt. 3, participants will be scanned as they watch a video version of the Raptor Center tour and during a subsequent set of Reactivation trials, and they will return one week later for a test session. However, during the reactivation phase of this experiment, a movie scene will be used to cue recall, followed by a fact presented on the screen. On 1/3 of the trials, the scene will be from the original video, and the facts will correspond to what was shown at that point; 1/3 of the trials will have an original video scene followed by a “lure” fact that were not shown in the video, and 1/3 of the trials will have video scenes that were not presented to the subject, along with a lure fact. During the final test, they will perform a recognition test consisting of studied facts and lure facts that were shown during the reactivation phase.

**Predictions:** SEM predicts that presentation of familiar video scenes during the reactivation phase will drive hippocampal pattern completion, leading to a spike in hippocampal activation and reinstatement of voxel activity patterns elicited during the corresponding point in the video (**H7**). SEM also predicts that, when familiar scenes are followed by new facts, hippocampal pattern completion should cause lure facts to be incorporated into PM network representations of the original video event (**H10**). Thus, on the final test, we expect higher rates of false recognition and higher PM network activation for lure facts that were preceded by studied scenes during the reactivation phase than for lure facts that were previously paired with new movie scenes.

**Experiment 5: Testing the prediction that hippocampal pattern completion can strengthen associations between scenes that are linked within the same event schema (**H9**), even if learning was distributed across different experiences**. **Procedure:** Participants in two experimental conditions will be scanned as they study sets of sequentially-presented sentences on the screen. Each set of sentences presents a narrative story about a primary character (“Liz”, “Kim”, etc.) and also includes a one sentence sideplot about a separate character, either “Jerry”

or “George”. Critically, across stories, the “Jerry” sentences will add up to a coherent sideplot (e.g., “Jerry told Liz, ‘I’m writing a movie script and I’ve sent it to Paramount Studios.’”; “Jerry told Kim, ‘Remember that script I wrote? I got a call from Paramount and they want to make it into a movie.’ ...etc.”) that can be accommodated within a single event schema, whereas the “George” sideplot sentences refer to multiple sideplot events that will be encoded within different schemas (e.g., “George told Liz, ‘I started taking an online class last week.’”; “George told Kim, ‘Starting tomorrow, I am going on a diet!’”). Stories will be normed in piloting studies such that they are matched for total number of event boundaries. 12 hours later, participants will return to the lab and they will be scanned during a recognition test that involves studied sentences from the sideplots, along with schema-consistent lure sentences matched to each studied sentence (e.g., “Starting tomorrow, I am going to exercise!”). For participants in the Sleep group, the study phase will be performed in the evening, and the test phase the next morning. The No Sleep group will perform the study phase in the morning and the test phase 12 hours later in the evening. (As a control for circadian effects, we will run a behavior-only Sleep group that performs the study phase in the evening and completes the test phase 24 hours later.) **Predictions:** We expect that retrieval of a sentence should also cue the hippocampus to pattern complete and reactivate the linked Jerry scene from the same event schema (**H7**), so we predict higher recognition accuracy and increased hippocampal activation for sentences from the coherent sideplot event compared to sentences drawn from the multiple sideplot events condition. Critically, because SEM is biased toward reconstructions that are schema consistent, the model predicts higher false alarm rates and increased hippocampal activation for lures that are consistent with the event schema for the coherent sideplot event than for lures related to multiple sideplot events sentences. We expect that participants in the Sleep group will benefit from off-line reactivation of the video events (**H11**), and therefore, we expect larger effect sizes for the Sleep than for the No Sleep group.

**Experiment 6: Testing the prediction that reactivation of event-specific information during sleep will enhance integration of linked event representations in the PM network. Procedure:** We will use the “targeted reactivation” technique introduced by Paller and colleagues<sup>88</sup>, in which auditory cues are used to manipulate reactivation of specific experiences during slow-wave sleep<sup>89</sup> (*Note that Norman’s lab is successfully using this technique in current studies*). Phase 1 of the paradigm will be similar to Expt. 5, except that: (1) each visually-presented cue will be accompanied by background music that is presented via magnet-compatible headphones, and (2) both “Jerry” and “George” sentences will add up to coherent sideplot events (i.e., we expect one event schema for each character). After the Phase 1, participants will be brought to the EEG booth in the Ranganath lab to take a nap with concurrent EEG recordings. Following Rudoy et al.<sup>89</sup>, during slow-wave sleep, we will play song clips that were previously heard while reading sentences that involved one of the sideplots. Subjects will return to the scanner for Phase 2, where they will be re-exposed to the same stories with the same musical accompaniment. Within PM network ROIs, we will examine similarity of voxel activity patterns elicited during exposure to all of the events depicted within Phase 1 and all of the events depicted within Phase 2.

**Predictions:** We expect that reactivation of the sideplot during the nap will alter voxel pattern similarity (PS) in the PM network during processing of sideplot sentences in Phase 2. For instance, if “Jerry” sentences were cued during the nap, we expect Jerry sentences to elicit highly similar voxel activity patterns (due to reactivation of a strong event schema) during Phase 2, as compared with Phase 1; this effect should be attenuated for the “George” events that were not cued during sleep. Note that this result could not be explained by standard consolidation theories,

which predict that reactivation simply strengthens memories or makes them less vulnerable to interference (cf. Muller & Pilzecker, 1900).

#### **Task 4: Specify the Cognitive and Neural Causes and Consequences of Event Segmentation** (J. Zacks, Washington University)

The overarching goal of this task is to use the SEM framework to account for human event segmentation at scale. To this end, two major questions need to be answered: (1) *Which computational and neural error signals lead to event segmentation?* (2) *What is the role of involuntary episodic retrieval in driving ongoing event segmentation and comprehension?*

##### **Subtask 1: Computational and Neural Error Signals that Lead to Event Segmentation.**

SEM proposes that *prediction errors* drive event segmentation, possibly relayed by the midbrain dopamine system (MDS) or the locus coeruleus norepinephrine system (LC) (**H1**). However, there are two alternative possibilities – *feature change* and *entropy*. In naturalistic activity, feature change, entropy, and prediction error are correlated. For example, when one walks into a party, many features of the situation are *changing* including location, characters, and objects. Walking through the door, one is likely relatively uncertain about exactly what will happen next (*high entropy*). And, whichever particular conversation or party game is predicted to occur, that particular thing is likely to result in a *prediction error*. Further, each of these three error signals could be computed in a global, modality-independent fashion, or locally within modality-specific processing systems. Here, we will address these two questions: 1) Does event segmentation depend on spikes in feature change, entropy, or prediction error? 2) Does segmentation depend on a unified, modality-independent signal or can a spike in multiple modality-specific signals lead to event model updating? To answer these, we will fit variants of SEM embodying these different assumptions to large event corpora, and then compare the models' timecourses of prediction error, entropy, and feature change to human behavioral and neurophysiological data.

**Stimulus materials.** We will construct a corpus of everyday activities performed in multiple locations by multiple actors using a large number of props. Each actor will perform multiple versions of a simulated morning consisting of several larger activities (*chapters*). The corpus will have naturalistic combinatorial structure, such that objects appear in multiple steps, steps appear in multiple tasks, and tasks

appear in multiple chapters (**Fig. 5**). Simulated mornings (n=4) will consist of 2-4 chapters, filmed as one continuous take, using 6 actors (3 female, varying ethnicity) for a total of 24 simulated mornings.

**Fig. 5.** Example components of videos in the activity corpus

Chapter	Task	Steps with objects
Making breakfast	Making coffee	Pouring water, adding coffee...
	Making tea	Boiling water, placing tea bag...
	...	
Working out	Bicep curls	Picking up floor mat, picking up dumbbell...
	Tricep dips	Moving chair, dipping...
	...	
Cleaning a room	Making the bed	Picking up pillow, placing it in pillowcase...
	Sweeping the floor	Opening closet, picking up broom...
	...	

##### **Experiment 1: Testing the computational basis of event segmentation.**

We will use Amazon Mechanical Turk to collect a robust sample of behavioral segmentation data for each of the 24 simulated mornings (100 participants/stimulus = 2400). Each participant will watch one simulated morning and segment it to identify event boundaries. We will train SEM models using cross-validation and then use the trained models to predict the distribution of segmentation judgments. First, we will test which potential segmentation signal predicts segmentation. Three models will be compared, which vary only in their segmentation mechanism: segmenting at spikes in (a) prediction error (**H1**), (b)

entropy, and (c) feature change. Although we expect prediction error, entropy, and feature change to be correlated, with this large sample we can use regression models to estimate the unique contribution of each signal to behavioral event segmentation. Second, using the best-fitting model, we will calculate the segmentation signal (prediction error, entropy, or feature change) separately for features related to entities, objects, and spatiotemporal framework as well as the overall signal. We will then test whether a single, unified segmentation signal is sufficient to account for human segmentation or whether the individual signals better fit the segmentation data.

**Experiment 2: Testing whether human predictions correspond to predictions generated by SEM.** SEM assumes that the ability to predict an unfolding event is the result of an inference based on previously experienced events. We will test this hypothesis by directly comparing human prediction performance to that of the model. Participants ( $n = 240$ , Mechanical Turk) will perform the *event prediction task*<sup>90</sup> with the same stimuli as in Experiment 1. In this task, the movie is stopped and two pictures are presented – one from 5 seconds later in the movie, the other from one of the actor’s other simulated mornings. Participants are instructed to select the picture that depicts what will happen in 5 seconds. We will select time points where the model’s performance is poor, average, or good (as indexed by prediction error). We will test the extent to which human observers’ prediction performance tracks that of the model.

**Experiment 3: Testing whether human anticipatory viewing behavior corresponds to prediction accuracy in SEM.** Research in infants and in adults performing everyday actions shows that humans make anticipatory eye movements based on predictions of future occurrences<sup>91</sup>, and we can leverage this finding by using eye movements to provide a continuous assay of prediction error. We will select a subset of 6 simulated mornings (1/actor) and code the locations in time and space of objects that the actor contacts. Each participant ( $n = 50$ ) will view 2 simulated mornings with instructions to watch for comprehension. We hypothesize that when the SEM’s error is low, the eyes will reach the to-be-contacted objects early; when the model’s error is high, the eyes will be later to arrive.

**Experiment 4: Testing neural mechanisms of event comprehension with functional MRI.** In this large-scale fMRI experiment ( $n=40$ ), participants will be scanned while viewing two simulated mornings. We will address multiple questions:

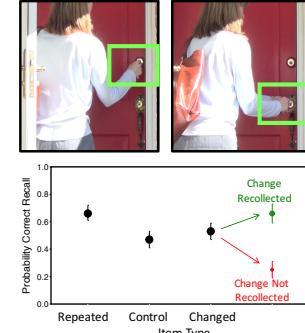
- 1) Does activity in the MDS, the LC, both, or some other system best track the error signal that predicts behavioral segmentation? We have previously shown that MDS activity tracks prediction error in everyday events<sup>90</sup> but in that study prediction error was necessarily confounded with feature change and entropy. We will use model-based fMRI to test whether prediction error, entropy, or global feature change as estimated by trained SEM models predict the dynamics of activity in the MDS, LC and other candidate structures (**H1**).
- 2) Does segmentation depend on a global error signal calculation for event perception (**H1**)? This is a reasonable hypothesis supported by some preliminary data. However, a strong alternative is that such signals are localized to domain-specific perceptual systems – predictions about object identity monitored within the ventral object-properties system, predictions about others’ movements monitored within the dorsal motor perception/control system, and so forth. To adjudicate between these, we will calculate the model’s error separately for features related to entities, objects, and spatiotemporal framework as well as the overall error signal. We hypothesize that medial PFC regions will track the integrated signal above and beyond the domain-specific

signals, and that spikes in activity in these regions will predict spikes in MDS (or possibly LC) activity, which in turn will predict behavioral segmentation.

**3)** Does the PM network represent information about the current event? Evidence from memory retrieval experiments and from studies of episodic future thought supports this hypothesis<sup>3</sup>, but alternatively, event representations have been hypothesized to depend on lateral PFC<sup>92</sup>. We will use RSA to test whether PM network regions and/or lateral PFC represent event models. Specifically, we will contrast PS between timepoints that span an event boundary identified by SEM against PS between timepoints that do not span an event boundary. We will also fit an HMM to the fMRI data<sup>27</sup> and use it to identify neural states in the PM network and lateral PFC that correspond to events in the film.

**Subtask 2: Testing the Role of Involuntary Episodic Retrieval in Driving Ongoing Event Segmentation and Comprehension.** We will use a novel experimental paradigm<sup>93</sup> (Fig. 6) to determine the role of hippocampally mediated episodic memory in ongoing comprehension. Subjects will watch a movie (Day 1) that includes several actions (e.g., feeding a dog) from one person's typical day. Next, they will view a second movie (Day 2), which includes actions that are *repeated*, *changed* (Fig. 6, top), and *control* actions not shown in Day 1. Changed actions involve performing the same goal (e.g., unlocking a door) using a different object (e.g., doorknob vs. deadbolt). We hypothesize that: **a)** During Day 2 viewing, people will retrieve the previously-seen action from the Day 1 video (H7). **b)** Changed actions will elicit a spike in prediction error during Day 2 viewing. **c)** This elevated prediction error will drive formation of a new episodic memory (H6). **d)** Based on a-c, we expect that the magnitude of episodic retrieval from Day 1 during Day 2 viewing will predict subsequent memory for changed Day 2 activities. **e)** Additionally, elevated prediction error caused by retrieval of changed activities should lead to event boundaries (H1). Preliminary data in support of a, c & e are presented in Figure 6. Participants viewed the Day 1 and Day 2 movies, and then memory for Day 2 was tested after a one-week delay using cued recall. Participants were also asked whether they recollect what had changed in each action. As expected, repeated actions were remembered better than control actions. Changed actions either showed interference or facilitation, *depending on whether the change was recollected*. This is consistent with the view that when participants made a prediction error, this triggered computations that facilitate long-term memory (H6).

**Fig. 6. Top:** Example of changed activity. **Bottom:** Memory for Day 2 events. Colored points indicate memory for changed Day 2 events, as a function of whether participants recollected what changed from Day 1 (green) or not (red). Error bars: 95% CIs.



**Experiment 5: Test the hypothesis that event change produces explicit prediction errors.** We will use the event prediction task used in Expt. 1. Participants ( $n = 60$ ) will first view one simulated day (Day 1), and then view a second day (Day 2) including repeated, changed and control activities while performing the event prediction task. Each clip will be stopped just as the change is becoming evident. At this point, there is information present in the video to make an accurate prediction, but if one predicts based on the previous episode one will make an error. Finally, participants will complete a cued recall test in which we describe the beginning of each activity and ask them about the potentially changed feature. We hypothesize that event changes will lead to prediction errors, and that prediction errors will be associated with *better* subsequent memory. This counterintuitive prediction arises because the error spike leads to better encoding.

**Experiment 6: Test the hypothesis that event change produces errors in predictive looking.**

Participants ( $n = 50$ ) will first watch a simulated Day 1. They then will watch a simulated Day 2 while eye movements are recorded. We hypothesize that changed trials will lead participants to look toward the location of the previous object before looking to the correct location. Finally, they will complete a cued recall test for the potentially changed features on Day 2. We hypothesize that on trials with larger eye movement prediction errors, memory will be *better*, again because the error spike leads to better encoding.

**Experiment 7: Test the hypothesis that change-induced prediction errors lead to event boundaries.**

This experiment will parallel Experiment 6, except that on Day 2 participants will segment the video rather than have their eyes tracked. We will test the hypothesis that changed trials lead to the perception of an event boundary at the point of change (H1). We also will use the cued recall task to test the hypothesis that the elevated prediction error on changed trials will result in improved episodic memory (bullet point d above).

**Experiment 8: Test effects of change-induced prediction errors on cortico-hippocampal interactions.**

Participants ( $n = 40$ ) will first view a Day 1 movie during scanning. After an intermediate delay, they will be scanned while viewing a Day 2 movie. This experiment will test our hypothesis that hippocampal pattern completion leads to predictions that are violated in the Day 2 movie, thereby resulting in a large prediction error and triggering the formation of a new memory trace (H6). To index retrieval, we will briefly pause the Day 2 movie right before event changes occur, and we will compute the similarity of hippocampal voxel patterns present during the pause interval to patterns present during viewing of corresponding scenes in the Day 1 movie. We hypothesize that higher pattern similarity during the pauses (indicating pattern completion) will be correlated with the magnitude of neural prediction error signals in the MDS, and that larger prediction errors will lead to better subsequent memory for the changed feature.

**Task 5: Determine How Broadband and Oscillatory Activity Contribute to Event Segmentation and Episodic Memory** (O. Devinsky, NYU; U. Hasson and K. Norman, Princeton; C. Ranganath, UC Davis; J. Zacks, Washington University)

The fMRI methods used in Tasks 2-4 are well-suited for addressing questions about slow-timescale neural processes related to the construction and maintenance of event models and/or retrieval of episodic memory. In Task 5, ECoG recordings will be used to test hypotheses that require a high degree temporal resolution: (1) We will test precise hypotheses about the temporal order of neural events predicted by SEM by measuring changes in high-frequency broadband activity (HFB; 70-200 Hz). HFB measures have high temporal resolution, because they are tightly linked to local population-level neural activity<sup>94-96</sup>. (2) We will explore whether slow and fast neural oscillations – rhythmic, synchronous changes in the excitability of large neural populations – play a special role in coordinating event cognition. Several models have proposed a central role for theta oscillations (4-8 Hz) in sequence processing and prediction<sup>97,98</sup>. These ideas motivate us to determine whether a neurally-based implementation of SEM (see Task 1) will need to incorporate mechanisms for oscillatory synchrony. Data will be collected at NYU, and interpretation, analysis, and manuscript preparation will be done in close collaboration with Hasson, Ranganath, Norman, and Zacks.

**Data acquisition and analysis:** Electrode montages vary across patients based on clinical criteria, but typical montages include depth electrode contacts in hippocampus and temporal lobe and subdural grids or strips of platinum electrodes (2.3 mm diameter) placed in prefrontal, lateral

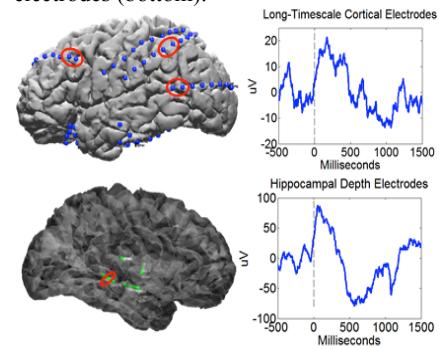
parietal, and lateral and ventral temporal cortex. Prior to data analysis, channels with epileptiform activity or excessive noise are identified and excluded. For the remaining clean channels, the raw time series is notch filtered at 60 Hz and its harmonics to remove power line noise. Data are then re-referenced to the common average of all clean channels. Signals were sampled at 30 kHz using a custom-built digital acquisition system<sup>99</sup>. Measures of instantaneous power and phase will be estimated for each band using a Hilbert transform<sup>100</sup>. We will focus analyses on power changes, in relation to task variables and in relation to variables generated by SEM. We will also test for phase reset (indexed by intertrial coherence) and between-condition changes in cross-frequency coupling. Nonparametric permutation tests will be used to assess statistical significance<sup>101</sup>.

**Experimental paradigm:** We expect to recruit 5 patients/year who are undergoing presurgical evaluation for intractable epilepsy. For maximal efficiency, recordings from sessions on two consecutive days will be used in a paradigm that will allow us to test several hypotheses about event representation. During Session 1, participants will watch a set of short (5 minute) “Coffee Shop World” videos used in Task 2. SEM will be used to calculate moment-by-moment estimates of prediction error, entropy, and feature change for these films. Recordings will be made continuously throughout the film and also during sleep. The next day, during Session 2, participants will see 3 kinds of videos: (a) “partial videos” corresponding to videos from Session 1, but with gaps. *Within-Event Deletions* will occur 10s after the onset of a new event, and then the screen will remain blank until 10s before the next event boundary. *Between-Event Deletions* will span an event boundary, such that the screen goes blank in the middle of one event it remains blank until the middle of the next event.

**Predictions:** During Session 1, we hypothesize that prediction error at an event boundary triggers the sending of cortical event information to the hippocampus (**H6**). Our pilot data (**Fig. 7**) indicate that prediction errors during viewing of film clips elicit large HFB spikes in the neocortex and hippocampus. We predict that the latency of HFB spikes at event boundaries will be earlier at neocortical sites than in the hippocampus. SEM-generated moment-by-moment estimates of prediction error, entropy, and feature change will be used in a regression analysis to test whether hippocampal and neocortical HFB spikes are related to prediction error (**H1**), or if they reflect entropy or feature change. We will test the latency of these spikes in order to test whether prediction error signals in PM network sites precede HFB activity increases related to memory encoding in the hippocampus. Finally, as in our fMRI analyses, we will use the HMM to identify neural representations of events and event boundary points in Session 1 films. Our first analyses will fit the HMM to the spatial pattern of HFB power across all sites that are free of epileptiform artifacts. We will also conduct analyses in which the HMM is fit to regional patterns of spectral power across hippocampal contacts and across PM network contacts.

In addition to investigating HFB power changes, we will conduct exploratory tests for cortico-hippocampal interactions during event boundaries using: (a) the weighted phase lag index (WPLI)<sup>102</sup>, which is a method of assessing oscillatory phase coherence that is insensitive to noise differences across experimental conditions<sup>102</sup>. In light of evidence showing that interregional theta phase coherence is increased during points

**Fig. 7.** Pilot data at NYU showing transient HFB power changes at interruption points between different movie clips at six cortical electrodes (top) and two hippocampal depth electrodes (bottom).



of prediction error<sup>103</sup>, we hypothesize that theta-band phase coherence between hippocampal and neocortical sites will be enhanced during event boundaries than during non-boundary points, even when regressing out effects of low-level visual features. (b) correlations in the envelope of HFB power, an index which is closely tied to fMRI intrinsic connectivity estimates<sup>104,105</sup>, (c) cross-regional phase amplitude coupling (cf. Voytek et al., 2015); for instance, during directed communication between the neocortex and hippocampus, it is possible that hippocampal gamma power (30-70 Hz) or HFB power will be modulated by theta phase (“phase-amplitude coupling”)<sup>100,106,107</sup>.

In Session 2, we hypothesize that, during partial videos, deleted scenes will lead to a high-entropy moment that should trigger hippocampal pattern completion (**H7**). Thus, during Within-Event Deletions, we should see increased HFB activity in the hippocampus *before* it is evident at neocortical sites (note that this is the opposite of what is predicted for Session 1). Additionally, hippocampal HFB power should be higher during Within-Event Deletions than during Between-Event Deletions, because hippocampal pattern completion should only facilitate retrieval of information within an event schema. Another way to test this hypothesis will be to use RSA to compute correlations between the pattern of HFB activity across distributed neocortical sites during the gaps and HFB patterns recorded as participants viewed corresponding time points in the film during Session 1. We will also use the HMM that was fit to the Session 1 movies to reconstruct neural states during the movie gaps. We expect that neural activity patterns during Session 1 will be reinstated during Within-Event Deletions, but the correlations will be markedly reduced during Between-Event Deletions. As noted earlier, theta oscillations, and especially theta-gamma phase-amplitude coupling<sup>97,98</sup> could play a role in supporting prediction of the scene sequences during the partial videos. To test this account, we will contrast theta power and within-site modulation of gamma power and HFB power by theta phase between Within-Event Deletions and Between-Event Deletions.

## **2. Potential Scientific Breakthroughs**

The proposed empirical research addresses major gaps in our understanding of human cognition. Neuroscience research on memory has largely been focused on hippocampal spatial representations during simple foraging behaviors, maze navigation, or fear conditioning tasks. Research in humans, in turn, has focused on memory for specific words or visual stimuli in lists. We know little about how the structure of complex, lifelike events affects memory and online processing. Given these major gaps in our current knowledge, even unanticipated outcomes would have a major scientific impact. By utilizing innovative data-driven machine learning analyses and theory-driven forward modeling of neuroimaging and ECoG data, every study in this project will break new ground.

Our research will be used to develop an entirely new computational framework that is scalable, meaningfully tied to neural mechanisms, and capable of representing structured relationships both within and across events. SEM is the first computational model to meaningfully account for neocortical representations of events, and it completely reconceptualizes the cortico-hippocampal interactions that support memory. Rather than dichotomizing semantic and episodic memory, SEM provides a framework for understanding how event knowledge shapes episodic memory encoding and how episodic retrieval accelerates the development of semantic event knowledge. Because schemas are fundamental to virtually every aspect of high-level cognition in the human brain, SEM will have a major impact on a wide range of areas, including episodic and semantic memory, reasoning, language

comprehension, planning, imagination, spatial navigation, and social cognition. Our development of a forward model of neural activity will allow us to make an unprecedented advance – to read out brain activity at a given moment and reconstruct cognitive representations of future events. It is difficult to overstate the potential impact for the scientific community or the relevance of this work to DoD capabilities.

### **3. Education, Training, and Broader Impacts**

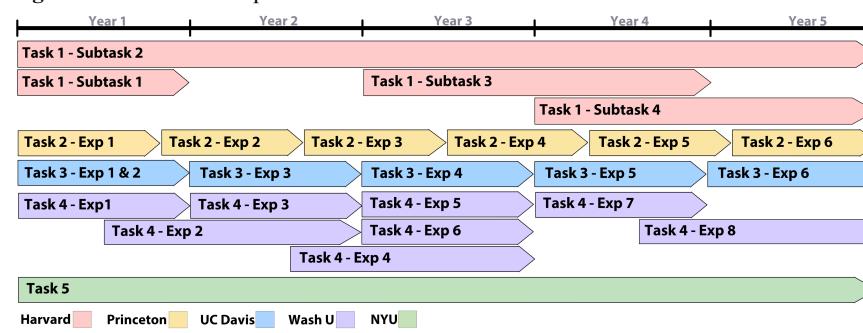
This 5-year MURI will contribute to the training of at least 10 postdoctoral researchers and 10 graduate students who will become proficient in the most advanced methodological and theoretical approaches in computational and cognitive neuroscience. Because we each supervise a large number of undergraduate research assistants, this project will contribute to the education of at least 100 undergraduates. Exchanges of trainees across project sites, regular videoconferences, and annual meetings (see Management Approach) will further enhance the educational impact of the project. The project's impact will extend beyond research personnel because each PI is heavily engaged in public outreach. For instance, Zacks has written a widely acclaimed book for lay audiences ("*Flicker: Your Brain on Movies*") and Hasson's TED talk has been seen by over 1,000,000 viewers. Given the broad relevance of our project, there is no doubt that the results will be of significant public interest.

### **4. Project Schedule, Milestones and Deliverables**

The empirical research involves novel and complex analyses of very large neural datasets, and the computational tasks address ambitious goals, but, as shown in Fig. 8, our timeline will allow us to complete these tasks. For Task 1, our first milestone will be a full-scale implementation of SEM by the end of Year 1. Other milestones include implementation of forward models of neural

activity for each of the empirical tasks as they are completed. By the end of Year 4, we anticipate completion of a biologically plausible SEM implementation. We expect that, for Tasks 2-5, new data collection should be completed for at least one experiment each year, so the

**Fig. 8.** Timeline for Completion



team's annual meeting (see below) will serve as a progress milestone, where each team will present results from at least one new study. By the end of the project, the team aims to deliver a version of SEM that can translate in any direction between pixel-level video stimuli, complex natural language stimuli, and brain activity patterns measured with ECoG and fMRI.

### **5. Management Approach**

The vision and scientific direction for the project is determined by the Leadership Team: Ranganath (PI), Zacks, Norman and Hasson, and Gershman (**Fig. 9**). Empirical fMRI and behavioral studies will be conducted at the UC Davis, Washington University, Princeton, and NYU sites. Orrin Devinsky will direct the NYU subcontract for completion of Task 5. Five co-investigators as well as Postdoctoral researchers, Graduate Students, and Undergraduates will

actively contribute, playing essential roles in theory development, experimental design and implementation, data acquisition, analysis, and interpretation.

## 5a. Interaction Management

Regular communication between team members across all levels is necessary to foster cohesion in this large-scale research endeavor. To meet these challenges, the team will communicate via:

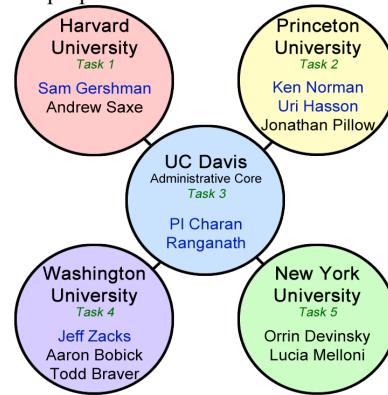
1. Video Conferences. The leadership team will continue their regular biweekly videoconference schedule. Meeting agendas include recent technical progress, prioritization of effort around upcoming deadlines (paper submissions, agency reporting, and deployment activities) and coordination of inter-site plans.
2. Inter-site Exchanges. Each year, a project site will host graduate or postdoctoral trainees from one or more other sites for extended (e.g., two-week) visits for methodological training, coordination of research protocols, and/or collaborative data analysis.
3. Collaborative Technologies. Each MURI team member and each external collaborator will be given a paid account on MemoryMURItteam.Slack.Com. The Slack platform enables users to organize interactions into separate “channels.” Each empirical study and modeling task will be given a separate channel to report progress, share data, and exchange ideas. Additional channels are used for meeting coordination, grant administration, exchanges of relevant journal articles, etc. All interactions are automatically backed up and archived. The team will also use shared Google Drive folders to enable collaborative manuscript preparation and share complex datasets. In keeping with confidentiality standards for human subjects research, no identifiable participant information will be posted on these platforms.
4. Website. Our website will be used for public dissemination of publications, conference presentations, data analysis code, annotated video stimulus files, etc. It will also include user-editable “Wiki” pages to disseminate knowledge and facilitate training of new personnel.
5. Annual Meeting. All MURI team members and DoD program staff attend an annual, in-person meeting to review progress, discuss new or unexpected research opportunities, and plan for the upcoming year. We will also invite an outside speaker to give a keynote address and provide advice on the direction of the project. The keynote address and team member presentations will be open to the media and general public.

## 5b. Collaborations/Subawards

**PI and Leadership Team:** Team member qualifications, along with their committed effort to the project, and other funded research activities are summarized below. Note that none of the funded projects overlap with the proposed research, and all investigators will have sufficient time to contribute the effort committed to the project. Committed effort, other funded research activities, and relationships to the proposed effort are summarized for the Leadership Team below (details of all funded and pending projects are summarized in the curriculum vitae):

Charan Ranganath (1.45 Months to MURI) will be the primary point of contact with ONR

**Fig. 9.** Organization of MURI Team.  
BLUE: Scientific Leadership Team.  
Each Project Site will assume responsibility for completing one of the proposed Tasks.



program chiefs and staff. He will direct the Leadership Team, coordinate activities across all project sites, oversee completion of Task 3, and direct analysis of ECoG data collected at NYU. Ranganath has a track record of leading large-scale, collaborative multimodal imaging projects on the cognitive neuroscience of human memory. Ranganath brings over 20 years of expertise in advanced fMRI, scalp EEG, and ECoG methods; his research on the neural representation of context in episodic memory<sup>16</sup> and his “Posterior Medial/Anterior Temporal” (PM/AT) model<sup>3</sup> was central to the anatomical framework and approach to modeling cortico-hippocampal interactions in SEM. Ranganath is funded by a DoD Vannevar Bush Faculty Fellowship [VBFF] (Decoding and Enhancing Neural Mechanisms for Episodic Memory, 02/01/15-01/30/20, 3.0 Month); NIH R01MH105411 (12/01/14-11/30/19, 1.5 AMs); and NIH EY025999 (4/01/16-03/31/21, .5 Month). Other commitments will be completed before or shortly after the start date. Funded effort is used to reduce Ranganath’s teaching commitments, which will allow sufficient time to lead the project.

Sam Gershman (*2 Month, year 1; 1 Month years 2-5*) is PI of Task 1, and he will lead all computational modeling activities. Gershman is an Early Stage Investigator, and his LCM framework plays a central role in the schema inference mechanism of SEM. He is currently funded by small private grants and awards (Total Other Effort=2 Months).

Ken Norman (*1 Month*) is PI of Task 2. Norman co-developed the CLS model (Norman, 2010; Norman & O'Reilly, 2003), a mechanistic, biologically-inspired architecture that will be used to model hippocampal function in this project. He is funded by a Templeton Foundation Program (3 Months); NIH R01MH069456 (09/19/16-06/30/21, 1.5 Months); NSF BCS1533511 (09/01/15-08/31/18, 1 Month); and NSF BCS1461088 (04/01/15-03/31/18, 0.4 Months). Norman is also co-PI on a grant from Intel Labs that will provide computational infrastructure for the team to optimize fMRI analysis methods with new, high-performance algorithms for machine learning and parallel computing.

Uri Hasson (*1 Month*) is co-PI of Task 2. He brings expertise in innovative fMRI methods and experience in studying neural responses to complex films, narratives, and other stimuli in realistic contexts. Hasson, Norman, and co-investigator Pillow developed the HMM analysis approach that will be used in multiple projects to identify stable neural representations in fMRI data. Hasson is funded by NIH 1DP1HD091948 (09/30/16-07/31/21, 2 Months). He also has two pending R01s (Neural dynamics supporting integration and recall over long timescales during natural continuous input, NIMH, 01/01/17-12/31/21, 2 Months; Brain-to-brain dynamical coupling: A new framework for the communication of social knowledge, NIMH, 03/01/17-02/28/22, 2 Months). The pending projects do not overlap with the proposed work.

Jeffrey Zacks (*1 Month*) is PI of Task 4. His pioneering research on event segmentation and his computational implementation of EST (Reynolds et al., 2007) laid the foundation for our approach to modeling of event schemas in SEM. Zacks' methods for identifying event boundaries will be used in all of the projects. He is currently funded by NIH R21AG052314 (09/01/16-08/30/18, 1 Month).

**ECOG (Task 5) Collaborators:** Dr. Orrin Devinsky (.6 Month) will direct the NYU subcontract (Task 5) and oversee patient care and recruitment at NYU Medical Center. During the project period, he will be funded by NIH and private grants with a total effort commitment of 9 Calendar Months. Dr. Lucia Melloni (.96 Month) organizes all intracranial EEG research at the NYU ECoG Center. Melloni will directly supervise the postdoctoral researcher at NYU. Devinsky and

Melloni have a track record of successful ECoG research, including collaborative work with Uri Hasson<sup>99</sup>.

**Computational Collaborators:** Dr. Aaron Bobick (.6 Month), Dean of the School of Engineering at Washington University, is a pioneer in computer vision and in vision systems for robots that navigate complex events. He will co-supervise a postdoctoral fellow at Washington University who will implement the computational models for Task 4. His other funded effort commitments will be completed soon after the start of our project. Dr. Jonathan Pillow (.16 Month) develops analytic tools to understand representations and computations carried out in the nervous system. He will contribute expertise in statistical analysis and machine learning approaches. His other funded effort commitments will be completed soon after the start of our project. Dr. Andrew Saxe (.1 Month) will contribute expertise in training and analyzing deep neural networks, both from machine learning and neuroscience perspectives. He is funded by a Harvard Swartz Postdoctoral Fellowship, which will cost share his effort for the MURI. Dr. Todd Braver (.6 Month) is a leader in the computational analysis of cognitive control processes and a close collaborator with Zacks on the RNN implementation of EST. He will consult with Zacks and Bobick on computational modeling for Task 4. During the project period, he will be funded by NIH and private grants with a total effort commitment of 2.6 Calendar Months.

### 5c. Facilities and Capital Equipment

MRI – At UC Davis, scanning will be done at the MRI Facility for Integrative Neurosciences (across the street from the Ranganath lab), with a 64-channel, 3 Tesla Siemens Skyra MRI System (Siemens Healthcare, Erlangen, Germany) equipped with a 32-channel head coil to support parallel and multiband imaging. Scanning at Princeton will be done at the Princeton Neuroscience Institute (PNI), using a 3 Tesla Siemens Prisma scanner that is equipped with 128 receive channels and a 64-channel head-neck coil. At Washington University, scanning will be conducted at the Research Imaging Center (RIC) of the Mallinckrodt Institute of Radiology using a Siemens Prisma MRI scanner with the same configuration as the scanner at PNI.

ECoG – Each year, approximately 50 patients undergo intracranial electrode monitoring and resective surgery at the Comprehensive Epilepsy Center (CEC) at New York University, making it one of the largest centers in the US to perform electrode implantations in the human brain. Each year, 30-40 patients are successfully recruited for participation in ECoG studies. NYU will recruit 5 patients/year for the proposed work, which is feasible based on their recruitment rates.

Computation – The PNI has two computational clusters that will be used in the project. The first is a 38-node, 912-core Xeon E5-2680v3-based platform, with 10GB RAM per core, and 10 Gb and FDR Infiniband interconnects. The second is an NSF-funded 50-node, 1000-core system, with 100 Intel Xeon Phi co-processors which provide 6000 additional Phi cores. At Washington University, we will use resources at the Center for High-Performance Computing, which offers access to ~1,800 computing cores, 10TB of memory and 19TFLOP/s of computing power as well as high-speed networking, bulk storage and technical support.

**6. Human Subjects:** Proposed human subject research will be carried out at UC Davis, Washington University, Princeton and NYU sites upon IRB approval at each site.

**7. Other Parties to Whom the Proposal Will Be Sent:** This proposal has not and will not be sent to any other funding agency until we receive a funding decision from ONR.