Distributed Patterns of Functional Connectivity Underlie Individual Differences in Long-Term Memory Forgetting

Yinan Xu[1,*], Chantel Prat[1,2], Florian Sense[3,4,5], Hedderik van Rijn[3,4,5], and Andrea Stocco[1,2]

[1] Department of Psychology, University of Washington, Seattle, WA USA

[2] Institute for Learning and Brain Sciences, University of Washington

[3] Department of Experimental Psychology, University of Groningen

[4] Behavioral and Cognitive Neuroscience, University of Groningen

[5] Department of Psychometrics and Statistics, University of Groningen

[*] Now at the Department of Psychology, University of Houston

Author Note

Correspondence should be sent to Andrea Stocco,

Campus Box 351525, University of Washington,

Seattle, WA 98195

Email: stocco@uw.edu

Abstract

Despite the importance of memories in everyday life and the progress made in understanding how they are encoded and retrieved, the neural processes by which declarative memories are maintained or forgotten remain elusive. Part of the problem is that it is empirically difficult to measure the rate at which memories fade and, without such a measure, it is hard to identify the corresponding neural correlates. This study addresses this problem using a combination of individual differences, model-based inferences, and resting-state functional connectivity. The individual-specific values of rate of forgetting in long-term memory (LTM) were estimated for 33 participants using a formal model fit to data from an adaptive fact learning task. Individual rates of forgetting were then used to examine participant-specific patterns of resting-state fMRI connectivity, using machine-learning techniques to identify the most predictive and generalizable features. Consistent with the existing literature, our results identified a sparse, distributed network of cortical and subcortical regions that underlies forgetting in LTM. Cross-validation showed that individual rates of forgetting were predicted with high accuracy (r = .96) from this connectivity pattern alone. These results open up new opportunities for the study of individual differences in LTM function and dysfunction.

Keywords: cognitive modeling, long-term memory, resting-state fMRI, functional connectivity

Introduction

The ability to acquire and maintain knowledge in long-term memory (LTM) is central to to learning, flexible behavior (Gershman & Daw, 2017), and even to the sense of personal identity (Klein & Nichols, 2012), and its loss is one of the most debilitating aspects of neurodegenerative diseases. Individual differences in this ability scale up to affect important real-life outcomes, such as differences in student grades (Sense et al., 2021) and ability to remember life-saving skills (Sense et al., 2019) Despite this, relatively little is known about individual differences in long-term memory. The temporal range of long-term memory, which spans from seconds to decades, makes it difficult to measure it in a laboratory setting. Consequently, previous studies have relied on individuals exhibiting dramatic differences in memory function, such as amnesic patients (such as patients suffering from medial temporal lobe injury: (Corkin, 2002; Gabrieli et al., 1988; Scoville & Milner, 1957)) or individuals with outstanding memory (such as individuals with Highly Superior Autobiographical Memory). Studies of individual variation in long-term memory across the normal population, instead, have restricted their investigation to specific, temporally circumscribed facets of long-term memory, such the moment of encoding (Alkire et al., 1998) and the moment of retrieval (Danker et al., 2008; Thompson-Schill et al., 1997).

Encoding and retrieval, however, are the two bookends of the life cycle of a memory. This paper, in contrast, focuses on the much less understood stage that happens between the two, during which a memory is consolidated or forgotten. While encoding and retrieval processes are largely centered around the medial temporal lobes and the lateral prefrontal cortices, respectively, consolidation and forgetting are distributed processes that involve large-scale changes in the connections between the medial temporal lobe and the entire neocortex (Alvarez

& Squire, 1994; Hardt et al., 2013; Moscovitch et al., 2005; Nadel et al., 2000; Squire, 2004), taking the form of synaptic changes, reorganization, and even neurogenesis (Frankland et al., 2013).

Thus, individual differences in forgetting must be ultimately underpinned by individual differences in the network of regions that is involved in consolidation and forgetting. Specifically, we hypothesized that individual differences in LTM function would relate to individual differences in resting state functional connectivity. Resting-state functional connectivity is typically measured with resting-state fMRI (rs-fMRI), continuous recordings of the spontaneous low-frequency fluctuations in blood oxygen level dependent signal while participants are awake but not actively engaging in any activities. The correlations between these spontaneous oscillations are systematic, with ensembles of voxels oscillating together (M. D. Fox et al., 2005) and forming functionally recognizable brain networks (Power et al., 2011; Yeo et al., 2011). Although the exact nature of spontaneous low-frequency activity is unknown, a dominant hypothesis is that it serves to maintain stable patterns of connectivity that encode Bayesian priors, shaped previous experiences, about the type of information that the brain is likely to encounter in the future (Pezzulo et al., 2021). This putative function of spontaneous brain activity, together with its distributed nature of, makes it an ideal substrate for LTM.

Resting-state activity is also reliable within participants (Braun et al., 2012). It has been speculated that it reflects the “architecture” of the brain (Stocco et al., 2021); small variations within this architecture are characteristic to every individual, to the point that it can be used to “fingerprint” individual participants (Finn et al., 2015; Waller et al., 2017). Consistent with this finding, individual differences in the resting-state functional connectivity between different regions have been shown to be predictive of performance differences in a variety of cognitive

abilities, including working memory (Avery et al., 2020), executive functions (Reineberg et al., 2015; Reineberg & Banich, 2016), motor learning (McGregor & Gribble, 2017), perceptual discrimination (Baldassarre et al., 2012), and most importantly, recognition memory (Persson et al., 2018). Thus, resting-state connectivity functional connectivity provides an ideal target to identify the biological underpinnings of individual differences in long-term memory retention and forgetting.

A Computational Theory of the Rate of Forgetting

To connect rs-fMRI measures to individual differences in forgetting, we need a way to estimate the degree to which individuals vary in forgetting. Purely behavioral measures of psychological constructs often have poor reliability, which leads, in turn, to poor brain-behavioral associations (Marek et al., 2020). One way to obviate this limitation is to fit realistic computational models to individual behaviors, and use the fitted model parameters as dependent variables (Collins, 2018; Daw, 2011). When the model captures the dynamics of the underlying neural process, its parameters can be more reliable and interpretable than pure behavioral measures. This has been shown, for example, with the threshold and drift parameters of accumulator models in perceptual decision-making (White et al., 2014, 2016) and in learning rate parameters in reinforcement learning models of learning (Xu & Stocco, 2021). Finally, using realistic models of long-term memory avoids the pitfalls that were originally pointed out by Loftus (1978; Wagenmakers et al., 2012) in applying statistical linear models to non-linear effects in memory performance.

In this study, the estimation of the rate of forgetting was based on a rational model of episodic memory (Anderson & Schooler, 1991; Pavlik & Anderson, 2008, 2005; Sense et al., 2018; Sense et al., 2016). This model, coupled with an adaptive memory task that minimizes the

use of alternative strategies, yields individual rates of forgetting for individuals that can be measured within a short time but can be reliably extrapolated over very long times (Sense et al., 2019).

In this model, the probability of retrieving a memory *m* at time *t* is proportional to its activation $A(m, t)$, a scalar quantity that reflects a memory's absolute availability. As in the multiple trace theory (Moscovitch et al., 2005; Nadel et al., 2000), each memory *m* is associated with multiple episodic *traces*, each of which corresponds to a specific event in which *m* was retrieved or (re)encoded, and each of which decays over time.

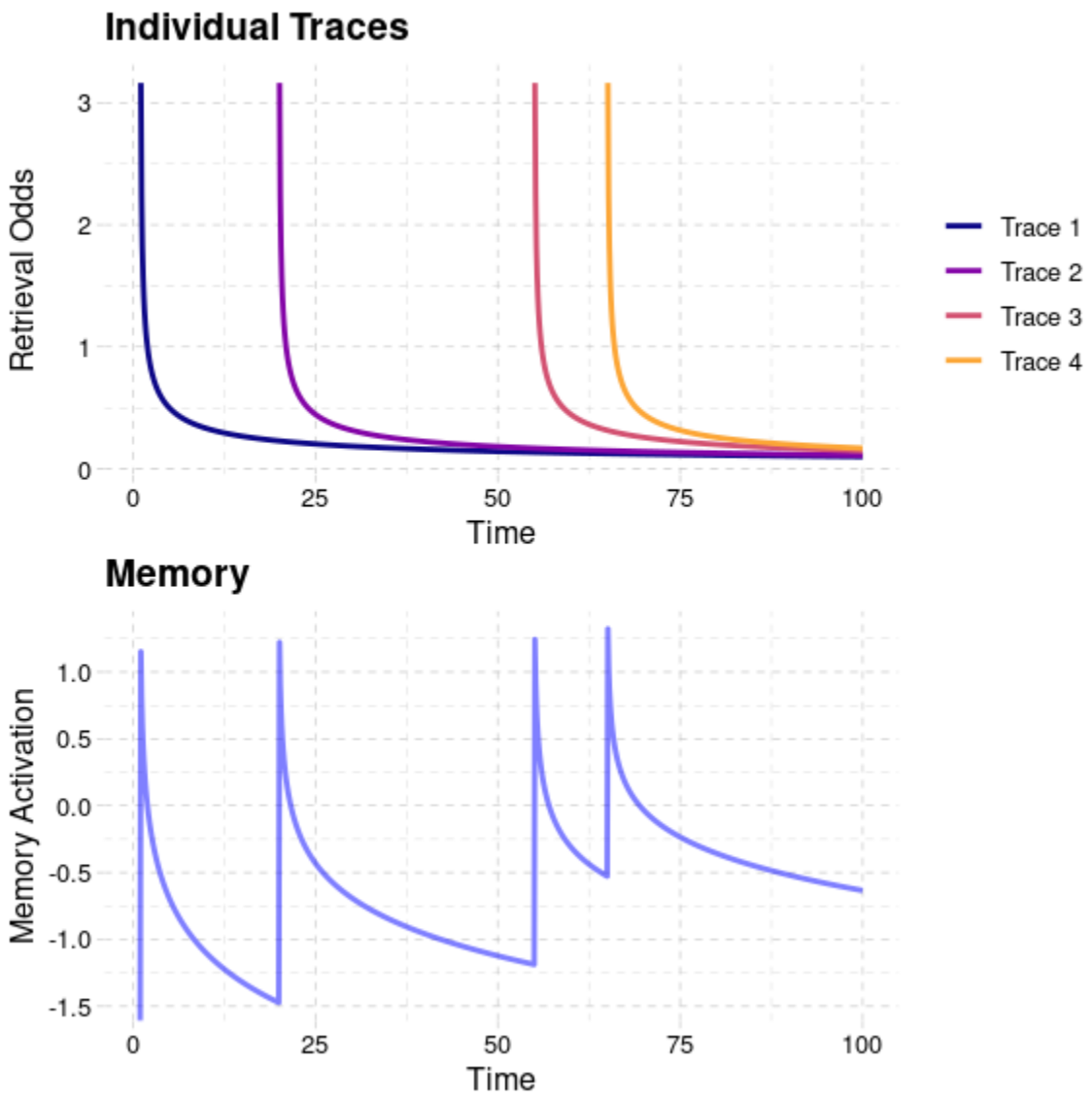

Figure 1: *Activation over time of a memory that has been encoded in four different events, at times t = 1, 20, 55, and 65 seconds (top). Each episode leaves an independent decaying trace (bottom). The total activation of the memory is the log sum of the decaying traces .*

The activation of each trace decays as a power function of time, and the activation of *m* is the log sum of the decaying activations of all its traces:

$$A(m, t) = \log \sum_i [t - t(i)]^{-d(i)}$$

where $t(i)$ and $d(i)$ are, respectively, the time of the creation of the *i*-the trace and its characteristic decay rate. A trace's decay rate, in turn, is related to the *m*'s activation at the moment each trace is created, thus capturing the fact that traces with higher initial activation decay faster than traces with lower activation (Pavlik & Anderson, 2005). Specifically:

$$d(i) = e^{A(m,\, t(i))} + \alpha$$

with $A(m, t(i))$ being the activation of *m* at the moment the *i*-th trace was encoded. Thus, if the times at which a memory *m* has been encoded, re-encoded, and retrieved are known and controlled for, its activation depends only on α, which is the only free parameter in the above equations and, by determining the decay of each memory, was taken to represent an individual-specific *rate of forgetting*.

Previous studies have shown that individual values of α are stable across time and materials and predictive of real-world outcomes such as learning and test scores (F. Sense et al., 2018, 2015; Florian Sense et al., 2016, 2019; van der Velde et al., 2020), thus meeting the psychometric criteria of reliability and validity (Shrout & Lane, 2012). Finally, at least one previous study has established that rates of forgetting, thus measured, reliably correlate with individual differences in neural activity (Zhou et al., 2021).

Experimental Predictions

To identify the neural correlates of these differences, this study used a statistical learning approach (James et al., 2013) in which a linear model was trained to predict an individual's rate of forgetting from their resting-state connectivity. If the rate of forgetting reflects a true biological parameter, individual variations in the rate of forgetting should reliably correlate with features of rs-fMRI and, in particular, with the distributed patterns of functional connectivity across networks that support memory consolidation.

Specifically, we hypothesized that the rate of forgetting would correlate with functional connectivity in the Default Mode Network (DMN: Raichle et al, 2001). The DMN encompasses most of the circuitry of memory encoding, including the hippocampus, the medial frontal cortex, and the precuneus (Ranganath & Ritchey, 2012). It is named so because it is active at rest but inactive while the brain is engaged in goal-oriented activity, a fact that has been explained as it being involved in rehearsing and maintaining memory as a form of "priors" for later processing (Pezzulo et al., 2021; Raichle & Snyder, 2007).

Memory consolidation also involves associative cortical areas in the frontal and parietal lobes (Alvarez & Squire, 1994; Durrant & Lewis, 2009; Moscovitch et al., 2005). These regions encompass multiple distinct networks, including the fronto-parietal network (FPN) and the cingulo-opercular network. Additionally, we expect that higher-level visual areas, such as those associated with the ventral and dorsal visual attention networks, will be involved as well, given their prominence in superior autobiographical memory (LePort et al., 2012; Palombo et al., 2018). Finally, because memory consolidation and decay occur as a system-wide re-organization of individual memory features, we expect that connectivity *within* these networks would be more important than connectivity *between* each of these networks.

In summary, we expect that the rate of forgetting would correlate with distributed patterns of connectivity between regions, and that these patterns would encompass the default mode, fronto-parietal, the cingulo-opercular, and visual attention networks. Furthermore, we hypothesized that measures of functional connectivity would be sufficient to precisely predict an individual's rate of forgetting. To test these predictions, we conducted the following experiment.

Materials and Methods

Participants

A total of 33 monolingual English-speaking participants (19 females) aged between 18 to 33 years old were recruited from a pool of University of Washington undergraduates for whom resting-state fMRI data had already been acquired in a previous, unrelated experiment. All participants provided informed consent and were compensated with a $25 gift card for their participation in the study. All of the recruitment and testing procedures were approved by the University of Washington's Institutional Review Board.

Vocabulary Learning Task

Rate of forgetting was estimated using the adaptive vocabulary learning task described in Sense et al. (2016). Twenty-five Swahili-English word pairs were selected from a previous study (Van den Broek et al., 2019). The task dynamically interleaved study trials and test trials. On study trials, a Swahili word (e. g., "samaki") and its corresponding English equivalent ("fish") were presented simultaneously on the screen. On test trials, only a Swahili word was presented on the screen as a cue (e.g., "samaki"), and participants were asked to respond by typing the corresponding English word ("fish") in an empty textbox. Trials were self-paced, with no cap on the time allowed for a response. Consecutive trials are separated by a 600 ms ISI after a correct response and a 4,000 ms ISI following an incorrect response during which feedback was shown

on screen. The order of repetitions and moment of introduction for each item were determined by the adaptive scheduling algorithm outlined below and described in more detail in van Rijn et al. (2009) and Sense et al. (2016). The task lasted approximately 12 minutes.

The presentation software was designed to optimally interleave the stimuli so that items would be repeated before the internal cognitive model predicted that they would be forgotten by continuously updating the rate of forgetting (α) for each pair. The cognitive model assumes that each memory *m* is a pair (see Anderson, 1974) that links a new Swahili word ("samaki") to a known English word ("fish"). In response to a study probe, participants use the presented Swahili word ("samaki") as a cue to retrieve *m* ("samaki" - "fish"). When *m* is retrieved, participants can type in the correct answer ("fish"). In turn, α is estimated from the time *T* it takes participants to respond (first key press), which is related to *m*'s activation by the equation:

$$T = e^{-A(m)} + k$$

where *k* represents a fixed offset (300 ms) for non-memory related processes (Anderson, 2007; Anderson et al., 2004).

rs-fMRI Data Collection and Preprocessing

A series of functional echo-planar images were collected. Each functional run lasted 7 minutes, and consisted of 210 brain volumes collected with TR = 2,000 ms, TE = 25 ms, FA = 72 degrees. Each volume consisted of 36 oblique 3-mm slices with no gap in-between, with an in-plane resolution of 80x80 3x3 mm voxels. Each functional series was preprocessed using a mixture of programs from the AFNI (Cox, 1996) and SPM (Penny et al., 2011) software packages. Specifically, each series of functional neuroimages was de-spiked, corrected for

differences in slice acquisition time, realigned to the mean image in the series, normalized the MNI ICBM152 template, and spatially smoothed with a 3D 8 mm FWHM Gaussian filter. In addition to the functional data, an anatomical MP-RAGE was also collected to aid in the normalization process. Functional connectivity data were analyzed with the Python Nilearn library (Abraham et al., 2014).

Brain Parcellation

To calculate functional connectivity, each participant's brain was divided into discrete regions using the parcellation method originally proposed by Power et al (2011). Although more recent parcellations have been proposed, this parcellation is notable for including both cortical and subcortical regions (see also Cole et al., 2016). Figure 2 depicts the coordinates of the 264 region centroids, color-coded according to their corresponding network, and using the color scheme proposed in the original paper.

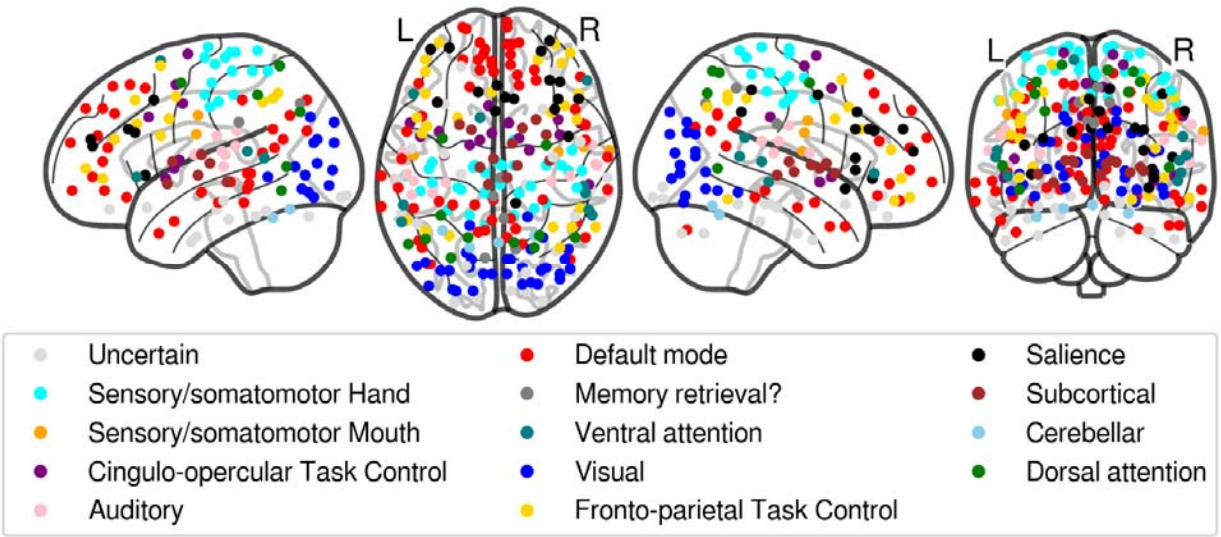

Figure 2. *Centroids of all 264 regions in the parcellation by Power et al. (2011) that was used in this study.*

Statistical Learning Model

A Lasso regression model was used to analyze the functional connectivity data. The choice was motivated by a number of considerations, which can be briefly summarized as follows. First, being a variant of linear regression, Lasso produces results that are relatively easy to interpret as linear weights of neural features; as such, these results are easy to evaluate and understand. Second, compared to other variants of linear regression (such as ridge regression), Lasso reduces the complexity of the model by evaluating whether coefficients can be shrunk to zero and, therefore, "drop" any unnecessary explanatory variables. In practice, this means that Lasso incorporates feature selection and parameter fitting in a single pass, further decreasing the experimenter's degree of freedom and reducing the dimensionality of the data.

Lasso is best explained as an extension of linear regression. In linear regression, a vector of parameters $\boldsymbol{\beta}$ is estimated to minimize the squared sum of the difference between the vector of observations $\boldsymbol{y}$ and the weighted explanatory variables $\boldsymbol{X}$, that is, argmin $||\mathbf{y} - \boldsymbol{\beta}\boldsymbol{X}||_2$ (where the notation $||\mathbf{v}||_n$ represents the L(n) norm of the vector $\mathbf{v}$). In Lasso, the difference is further penalized in proportion to the sum of the absolute values of the beta weights (their L1 norm):

$$\boldsymbol{\beta}^* = \text{argmin}(||\mathbf{y} - \boldsymbol{\beta}\boldsymbol{X}||_2 + \lambda||\boldsymbol{\beta}||_1)$$

The value of λ represents the tradeoff between model simplicity (encouraged by the $||\boldsymbol{\beta}||_1$ penalty) and accuracy (encouraged by ordinary least-squares minimization term $||\mathbf{y} - \boldsymbol{\beta}\boldsymbol{X}||_2$). When $\lambda = 0$, there is no penalty and the model reduces to canonical linear regression, privileging

fit at the expense of generalizability. As λ grows, however, so does the penalty for larger weights, until only the intercept remains, resulting in the simplest generalizable model.

A cross-validation procedure was used to find the optimal value of λ. A sequence of possible λ values was generated, and, for each value, the performance of the Lasso algorithm was measured using leave-one-out validation (LOOV). In LOOV, the algorithm is run 33 times, each time leaving out a different participant as the test set while the β values are fit on the remaining 32 participants as the training set. The mean error in predicting the rate of forgetting for the left-out participant was then measured for all values of λ, and the value of λ that produced the smallest cross-validation error across all participants was chosen. The optimal value corresponds to a minimum in the convex profile of the cross-validation error curve (G. James et al., 2013).

Results

Rates of Forgetting

Rates of forgetting were calculated for each individual by averaging across the forgetting rates for each word pair they studied. The resulting values ranged from 0.253 to 0.385, with a mean of 0.305 and standard deviation of 0.029. Both the mean and the range of values are in line with previously published studies (Sense et al., 2016; Zhou et al., 2021).

Connectome Extraction

Each participant's global pattern of brain connectivity, known as the *connectome* (Sporns, 2011), was extracted using a standard procedure (Cole et al., 2016; Shen et al., 2017). For each of the 264 x 264 pairs, the Pearson correlation coefficient between the respective timeseries was computed. The group-average of the resulting matrices (Figure 2A) shows a canonical distribution of *r* values, including the zones of higher *r* values along the diagonal (corresponding to connectivity within the same networks). This pattern is entirely consistent with

previous findings using the same parcellation scheme (e.g., Cole et al 2016).

One problem with this canonical connectome is that the raw correlation coefficients are almost all positive and partially driven by common sources, including unobserved physiological effects. To remove these confounds, we followed the procedure of Cole et al. (2016), and recalculated the correlation matrices using *partial* correlation coefficients, so that, when calculating the correlation between two regions, the correlations between each region in the pair and the remaining 262 regions were partialled out. Note that this procedure also eliminates any remaining residual physiological artifacts that are shared across brain regions. The resulting average connectome is a much sparser matrix (Figure 3B) that includes both negative and positive correlations (Cole et al., 2016; Fox et al., 2005).

Each participant's sparse correlation matrix was then reshaped into a row vector of (264 x 263) / 2 - 264 = 34,452 elements. To reduce the number of regressors, only connections between regions that exhibited a minimum average partial correlation coefficient of $|r| > 0.1$ across participants (see Figure 3B) were included. This threshold was chosen *a priori* to remove all of the functional connectivity values that, in our group of participants, would not be significant at a very liberal threshold of $p < .50$. The mean correlation coefficient across participants was calculated by first turning each individual coefficient into a *z*-value using Fisher's *r*-to-*z* transformation, then averaging the individual *z*-values, and transforming the average *z* value back into an equivalent *r* coefficient.

Thanks to this sparsification, only 301 connectivity measures per participant remained. Thus, the final regressor matrix $\boldsymbol{X}$ passed to the Lasso algorithm was a 33 (participants) x 301 (regressors) matrix.

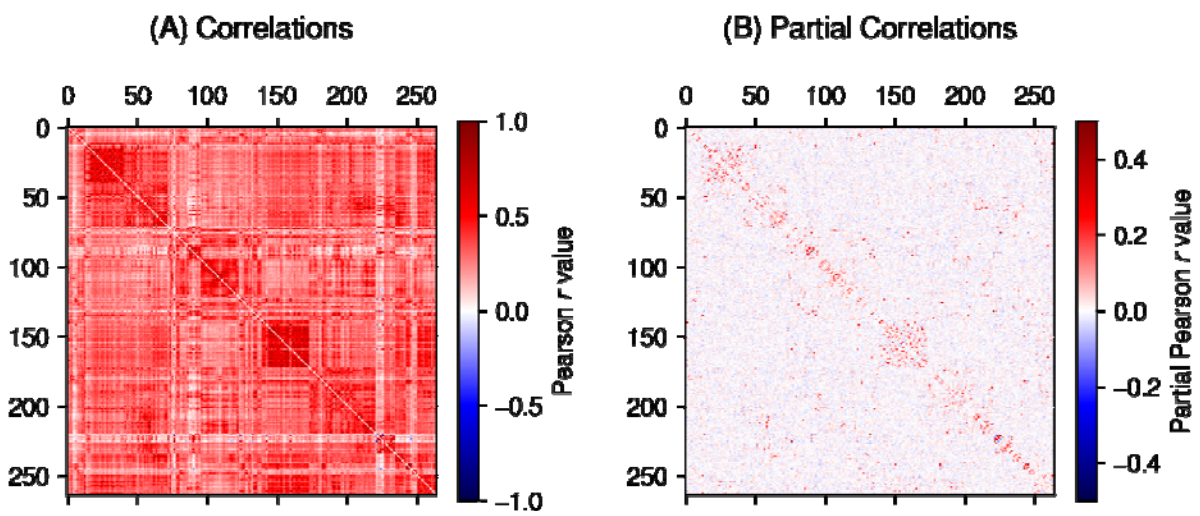

Figure 3. *(A) Mean raw correlations between each of the 264 regions across all participants (Figure 1); (B) Mean partial correlations between the same regions. In each matrix, rows and columns are ordered by network.*

Resulting Connectivity

Using the optimal lambda value obtained through LOOV (λ = 0.00152), we fit a single Lasso regression model with the 301 connections as input features to predict the rates of forgetting for 33 participants. Of the 301 initial features, only 25 were left with a $\beta \neq 0$, involving a total of 46 brain regions, with four of these regions being included in two connections and the other 42 belonging to a single connection. It was the individual variations in these connections that were found to be most predictive of individual rates of forgetting.Figure 4 shows which connections are relevant and which network each region belongs to. Individual ROIs are colored using the same convention of Power et al. (2011) and Figure 2. Edges between nodes are color-coded based on the corresponding β value estimated by Lasso; negative weights, in blue, are associated with smaller rates of forgetting and positive weights, in red, are associated with higher

rates of forgetting.

Consistent with our expectations, these connections were widespread and distributed, involving both the DMN as well as sensory and motor areas. Also consistent with the current views of synaptic reorganization during consolidation and forgetting, most of these connections (64%) are situated between pairs of regions *within* the same network, rather than *between* networks. This proportion is significantly different from what would be expected if the connections were randomly sampled from the global connectome, where within-network connections represent a minority of cases (9.67% of the total; $\chi^2(1) = 84.48$, $p < 0.0001$). Of the connections between regions belonging to different networks, 44% include regions in the DMN.

In addition to the DMN and primary and associative sensory areas, regions in the dorsal and ventral attention networks and in the striatum ("Subcortical" network in Power et al., 2011) were also associated with the rate of forgetting. Individual morphological differences in the striatum morphology in the activity of visual areas have been previously related to superior episodic memory (LePort et al., 2012; Palombo et al., 2018). The involvement of these regions might also be explained by the nature of the task, which uses visual word forms as retrieval cues, as both visual attention networks and the striatum have been implicated in directing visual attention (Stocco et al., 2017).

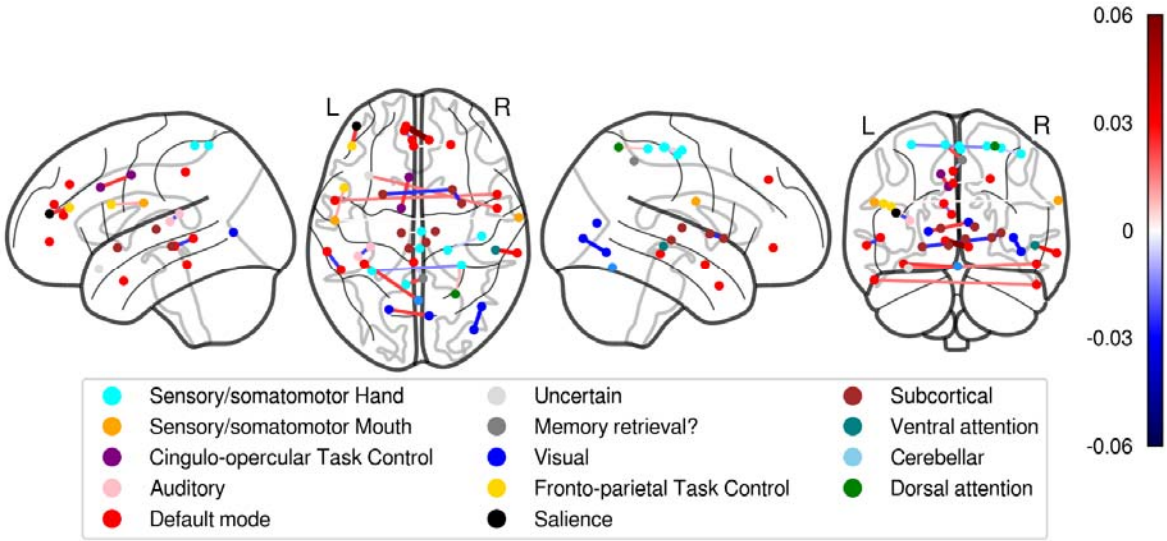

Figure 4: *Functional brain connections predictive of individual rates of forgetting. ROI centroids are colored according to the scheme by Power et al., 2011. The color of each connection reflects the value of the corresponding β weight.*

Validation of the Lasso Results

Although Lasso reduces the number of predictors to a number that is comparable to the number of observations, it has been argued that its performance is severely hindered when regressors are collinear, which is likely to happen in cases, such as this one, when the number of regressors is much larger than the number of participants (Jia & Yu, 2010). In such cases, given equally valid predictors, the penalty term might force the Lasso algorithm to select one of them at random, providing an inadequate and partial picture of which regressors are indeed essential. If this were the case, other equally valid predictors would have been overlooked. A simple test to exclude this possibility is to remove the 25 connections from the matrix of 301 sparse regressors $\boldsymbol{X}$, and re-run the Lasso procedure. If Lasso converges on a different and equally successful solution, then other important regressors were accidentally excluded.

When applied to our data, this test shows that the Lasso algorithm does *not* converge anymore. Specifically, the LOOV procedure fails to find an optimal minimum. Instead, the cross-validation error is much larger (up to 5x) than in the previous model and greater generalization (as indicated by smaller cross-validation errors) is progressively observed as the value of λ increases, until the penalty term is so large that only the intercept remains in the linear model. Figure 5 illustrates the cross-validation error profiles at different levels of λ for the original, full model (in blue) as well as the reduced model without the 25 predictive connections (in orange). The difference between the two curves confirms that these 25 connections are jointly essential for capturing the rate of forgetting.

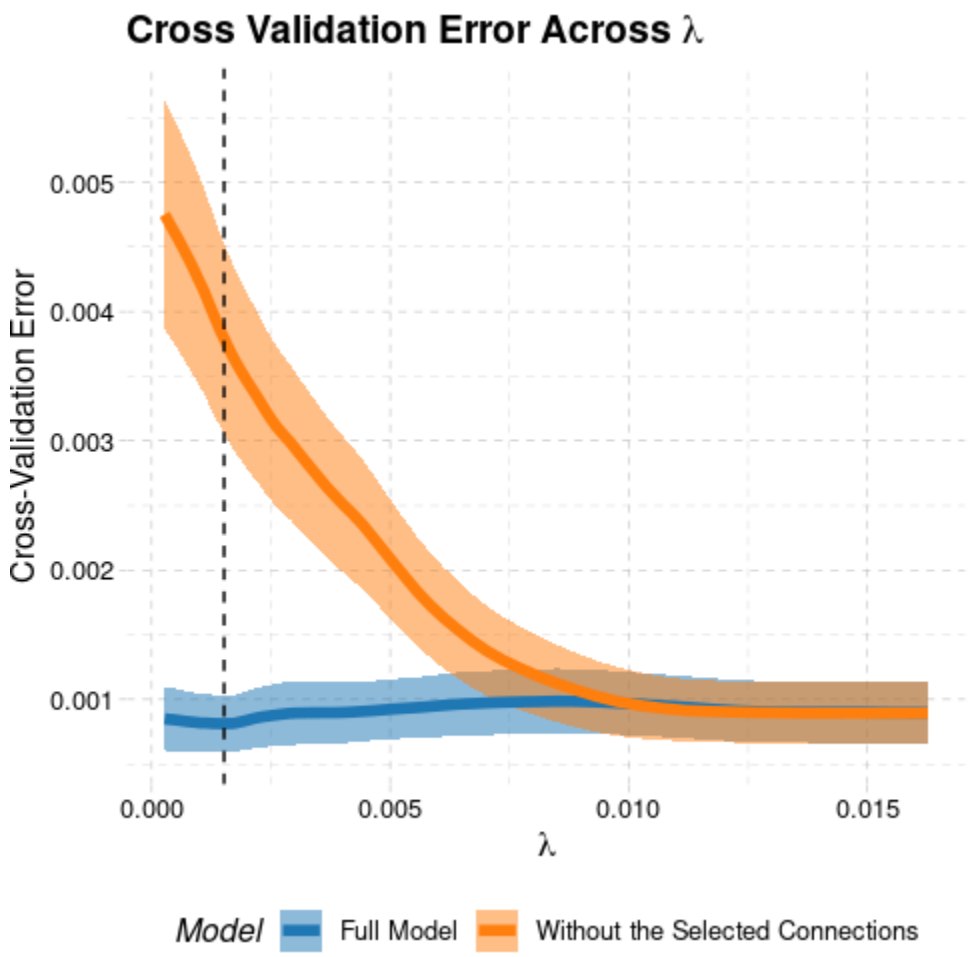

Figure 5*: Cross-validation error across different values of lambda for two different models. The blue line represents the profile of the "full" model containing all 301 connections; the black vertical line represents the value of λ that minimizes the cross-validation error; at that value, only 25 connections survive (see Figure 8 for their estimated β weights). The orange line represents the profile of a reduced model without the 25 connections identified at the optimal value of λ. For both curves, solid lines represent mean values and shaded ribbons represent standard errors.*

A second test consists of measuring the collinearity of the resulting model. Since the number of predictors in the model (p = 25) is comparable to the number of observations (n = 33), it is possible that the remaining predictors remain highly collinear, which would suggest that all connections in the brain contain similar information and are similarly related to the rate of forgetting. To exclude this hypothesis, the reduced model was examined using the Variance Inflation Factors (VIF) metric as implemented in R's *car* package (Fox & Weisberg, 2018). As a rule of thumb, VIF factors > 10 indicate collinearity (Fox & Monette, 1992). All of the remaining predictors had a VIF value < 10. In fact, the VIF values were < 5 in all but one case, which suggest that the remaining predictors contributed independently to an individual rate of forgetting.

Regional Importance Within the Network

The resulting connectivity matrix can also be analyzed in terms of network-level measures, such as the importance of a node (region) within the network. The importance of a node can be calculated as the L1 norm of all the edges connected to that node. In the case of our connectivity matrix, the L1 norm for a region is the sum of the absolute values of the beta weights of the connections that include that region. The importance of each region is visualized in Figure 6, where the 46 regions of Figure 4 are plotted with the size of the corresponding marker reflecting their importance. This visualization highlights the prominence of ventromedial prefrontal DMN regions and of the basal ganglia. This is consistent with one previous study (LePort et al., 2012), which similarly identified changes in the grey matter density of these two regions as being associated with superior autobiographical memory.

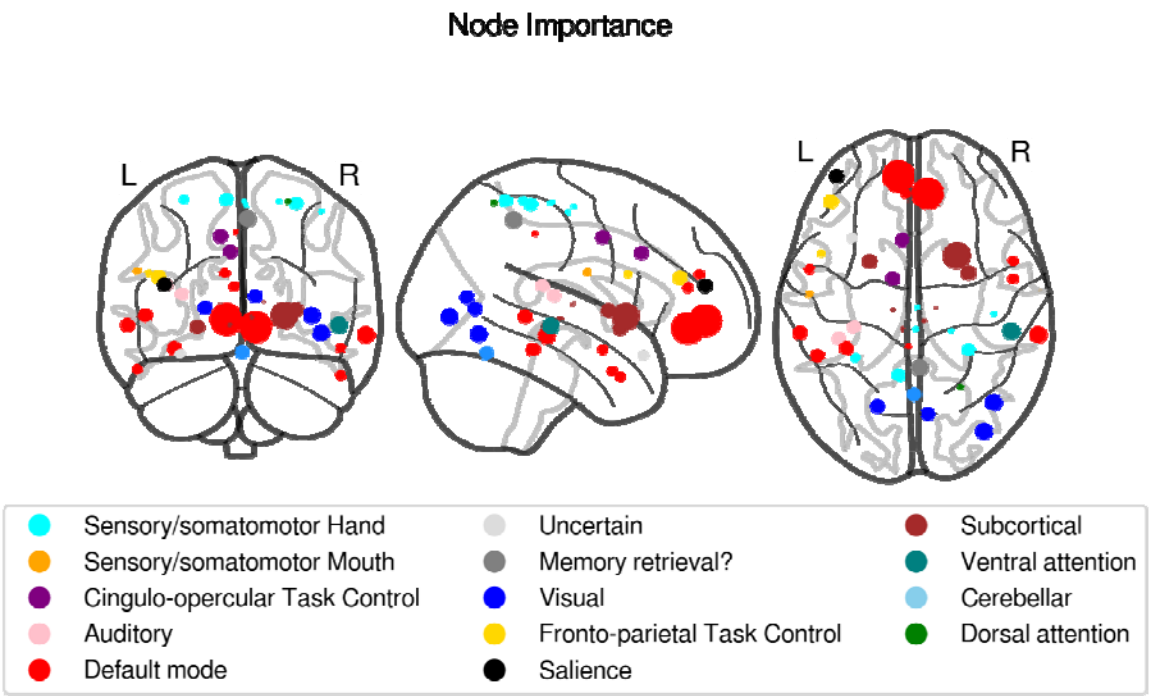

Figure 6: *Relative node importance of each region in the network of Figure 5. The size of each marker indicates the importance of the corresponding region, calculated as the L1 norm of its connections.*

Predicting Rate of Forgetting From Connectivity

One final question that remains to be addressed is how reliably the information from functional connectivity alone could be used to predict individual rates of forgetting. To do so, we used the LOOV cross-validation procedure, iteratively predicting the rate of forgetting of one test participant using their functional connectivity values and the beta weights estimated from the other 32 participants. The predicted rates of forgetting were almost perfectly correlated with the observed ones, with a Pearson correlation coefficient of $r(33) = .96$ ($p < 0.0001$) and an RMSE of only 0.008, or approximately 3% of the mean value of the true parameter (Figure 7). Thus, our results seem consistent and generalizable.

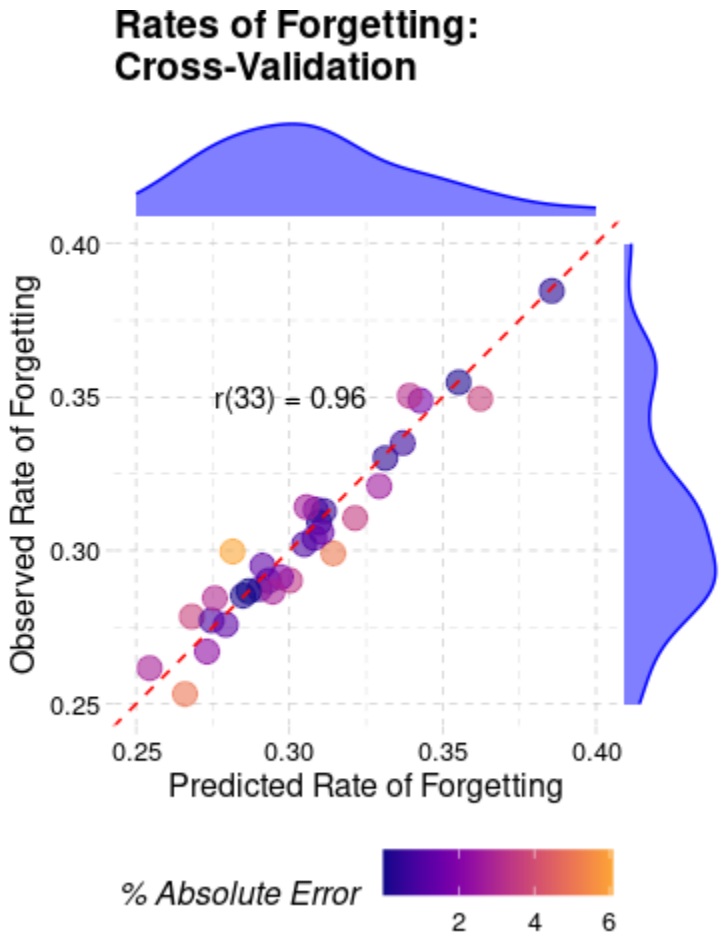

Figure 7: *Predicted vs. observed values of rate of forgetting across all 33 participants. Each dot represents a single participant; their color represents the percentage error in the estimate; the red dashed line is the identity line.*

In addition, we examined the stability of the β weights of each connectivity value obtained during the cross-validation procedure. This last measure is included because, given the small sample, it is possible that a given set of regressors could yield good predictions even if the specific weights of each regressor change dramatically across participants. Figure 8 depicts each connection's estimate across all participants sets as well as the corresponding 95% confidence

interval. As the figure shows, the estimated β weights were consistent across sets of participants, further supporting the idea that the corresponding connectivity values are reliable predictors of individual differences in long-term memory forgetting.

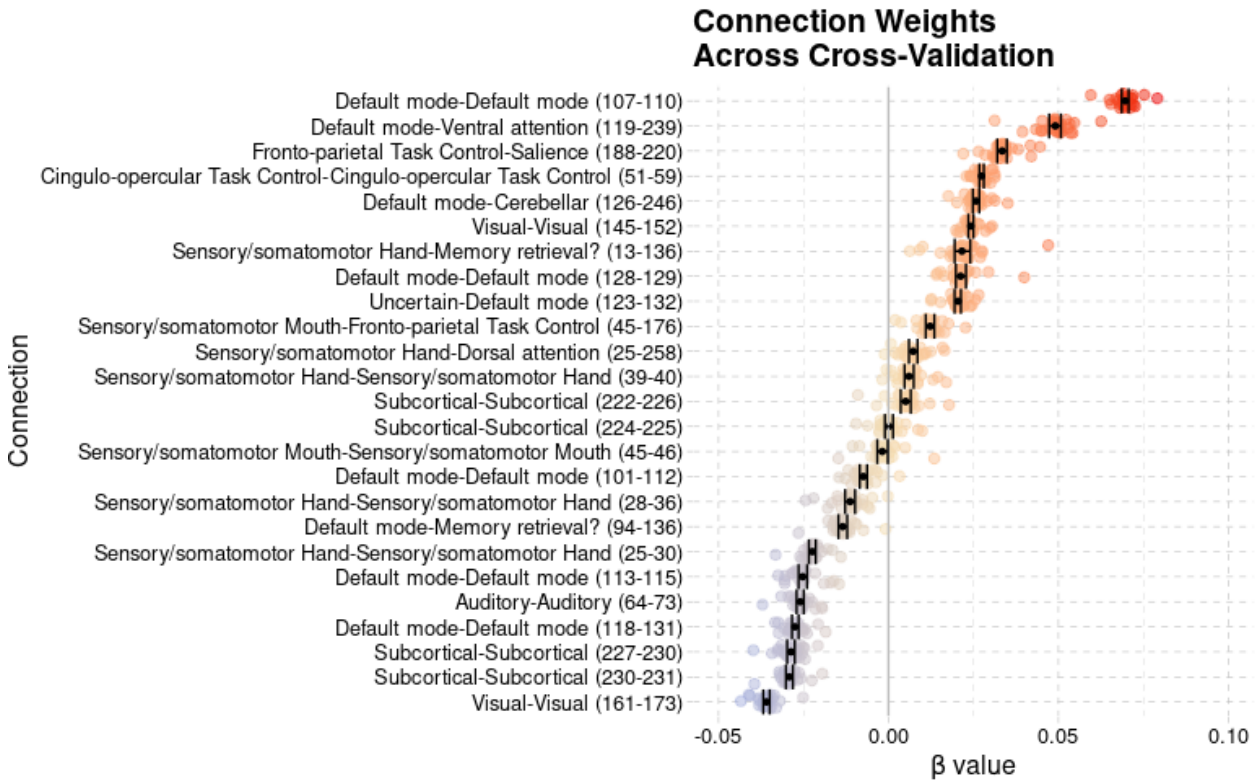

Figure 8: *Mean β values of each connection across all cross-validation cycles. Labels indicate the networks that are bridged by each connection; numbers in parentheses indicate the IDs of the regions bridged by each connection according to the Power et al. (2011) coding system; colored points represent individual estimates from a cross-validation run; black points represent means; and black error bars represent 95% confidence intervals.*

Discussion

In this paper, we provided evidence that an individual's rate of forgetting, operationalized as a decay intercept parameter in Pavlik and Anderson's (2005) model, can be predicted using aa distributed pattern of functional connectivity obtained from resting-state (or task-free) brain connectivity measures.This set of connections was identified through Lasso regression and cross-validation, and demonstrates remarkable generalizability and prediction accuracy while also reducing the full connectome to a manageable set of predictors.

While our results are encouraging, one limitation must be acknowledged, namely, that a small number of fairly homogeneous participants studied a relatively small set of foreign vocabulary items, which potentially limits external validity. This limitation could also explain the correlation coefficient of Figure 7, which is unusually large even if produced through cross-validation. Thus, future studies should expand the sample size and explore other sets of study materials. At the same time, these limitations also demonstrate that even in such a homogenous group, our approach can pick up on meaningful interindividual differences.

Importantly, our results are consistent with the existing literature. Specifically, the individual rate of forgetting correlates with the connectivity in networks associated with different facets of memory function, like the sensory and associative cortices (involved in representation) and the DMN (involved in memory encoding and retrieval). Additionally, our results show remarkable similarity to the EEG findings of Zhou et al. (2021). In their study, a significant positive correlation between power in the beta band and rate of forgetting α was observed distributing at bilateral prefrontal (channel AF3, AF4) and right parietal (channel P8) regions, in locations that are consistent with the ventromedial prefrontal DMN nodes and the dorsal visual attention nodes in Figure 6, respectively. Because greater EEG power is associated with greater

synchrony between neurons, it is not surprising that greater local functional connectivity in these regions was also positively correlated with rates of forgetting.

Finally, it is worth noting that, in addition to finding plausible connectivity correlates of the rate of forgetting, our findings can be applied in the opposite direction, to decode an individual's rate of forgetting from distributed patterns of connectivity using a standard brain parcellation. Such an approach could be used, for example, to detect the early onset of abnormal forgetting in individuals with known memory pathologies and cognitive degenerative diseases. Most importantly, this finding demonstrates the possibility of decoding model parameters directly from resting state data, rather than from a combination of specific tasks. If generalized beyond these results, this procedure could play an important role on the road towards developing high-fidelity idiographic models. This procedure also has the benefits of showing reflections of parameters in brain connectivity network regions, shedding light to a more holistic and comprehensive way of doing functional anatomy research on human cognition.

Acknowledgements

This work was supported by grant N00014-17-1-2607 from the Office of Naval Research to CP and by grant FA9550-19-1-0299 from the Air Force Office of Scientific Research to AS. The authors would like to thank Dr. Adel Lee for her helpful methodological comments and suggestions.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics, 8*, 14. https://doi.org/10.3389/fninf.2014.00014

Alkire, M. T., Haier, R. J., Fallon, J. H., & Cahill, L. (1998). Hippocampal, but not amygdala,

activity at encoding correlates with long-term, free recall of nonemotional information. *Proceedings of the National Academy of Sciences of the United States of America, 95*(24), 14506–14510. https://doi.org/10.1073/pnas.95.24.14506

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: a simple network model. *Proceedings of the National Academy of Sciences of the United States of America, 91*(15), 7041–7045. https://doi.org/10.1073/pnas.91.15.7041

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 6*(4), 451–474. https://doi.org/10.1016/0010-0285(74)90021-8

Anderson, J. R. (2007). *How Can the Mind Occur in the Physical Universe?* Oxford University Press.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036–1060. https://doi.org/10.1037/0033-295X.111.4.1036

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*(6), 396–408. https://doi.org/10.1111/j.1467-9280.1991.tb00174.x

Avery, E. W., Yoo, K., Rosenberg, M. D., Greene, A. S., Gao, S., Na, D. L., Scheinost, D., Constable, T. R., & Chun, M. M. (2020). Distributed Patterns of Functional Connectivity Predict Working Memory Performance in Novel Healthy and Memory-impaired Individuals. *Journal of Cognitive Neuroscience, 32*(2), 241–255. https://doi.org/10.1162/jocn_a_01487

Baldassarre, A., Lewis, C. M., Committeri, G., Snyder, A. Z., Romani, G. L., & Corbetta, M. (2012). Individual variability in functional connectivity predicts performance of a

perceptual task. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(9), 3516–3521. https://doi.org/10.1073/pnas.1113148109

Cole, M. W., Ito, T., Bassett, D. S., & Schultz, D. H. (2016). Activity flow over resting-state networks shapes cognitive task activations. *Nature Neuroscience*, *19*(12), 1718–1726. https://doi.org/10.1038/nn.4406

Collins, A. G. E. (2018). The Tortoise and the Hare: Interactions between Reinforcement Learning and Working Memory. *Journal of Cognitive Neuroscience*, *30*(10), 1422–1432. https://doi.org/10.1162/jocn_a_01238

Corkin, S. (2002). What's new with the amnesic patient H.M.? *Nature Reviews Neuroscience*, *3*(2), 153–160. https://doi.org/10.1038/nrn726

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, an International Journal*, *29*(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Danker, J. F., Gunn, P., & Anderson, J. R. (2008). A rational account of memory predicts left prefrontal activation during controlled retrieval. *Cerebral Cortex* , *18*(11), 2674–2685. https://doi.org/10.1093/cercor/bhn027

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. Decision Making, Affect, and Learning: *Attention and Performance XXIII*, 23, 3–38.

Durrant, S., & Lewis, P. A. (2009). Memory consolidation: tracking transfer with functional connectivity. *Current Biology: CB*, *19*(18), R860–R862. https://doi.org/10.1016/j.cub.2009.08.019

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., & Constable, R. T. (2015). Functional connectome fingerprinting: identifying

individuals using patterns of brain connectivity. *Nature Neuroscience*, *18*(11), 1664–1671. https://doi.org/10.1038/nn.4135

Fox, J., & Monette, G. (1992). Generalized Collinearity Diagnostics. *Journal of the American Statistical Association*, *87*(417), 178–183. https://doi.org/10.1080/01621459.1992.10475190

Fox, J., & Weisberg, S. (2018). *An R Companion to Applied Regression*. SAGE Publications.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9673–9678. https://doi.org/10.1073/pnas.0504136102

Frankland, P. W., Köhler, S., & Josselyn, S. A. (2013). Hippocampal neurogenesis and forgetting. *Trends in Neurosciences*, *36*(9), 497–503. https://doi.org/10.1016/j.tins.2013.05.002

Gabrieli, J. D., Cohen, N. J., & Corkin, S. (1988). The impaired learning of semantic knowledge following bilateral medial temporal-lobe resection. *Brain and Cognition*, *7*(2), 157–177. https://doi.org/10.1016/0278-2626(88)90027-9

Gershman, S. J., & Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annual Review of Psychology*, *68*, 101–128. https://doi.org/10.1146/annurev-psych-122414-033625

Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: the role of active forgetting in memory. *Trends in Cognitive Sciences*, *17*(3), 111–120. https://doi.org/10.1016/j.tics.2013.01.001

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). Statistical Learning. In G. James, D.

Witten, T. Hastie, & R. Tibshirani (Eds.), *An Introduction to Statistical Learning: with Applications in R* (pp. 15–57). Springer New York.

Jia, J., & Yu, B. (2010). On model selection consistency of the Elastic Net when *p* □ *n*. *Statistica Sinica*, 20(2), 595–611. https://www.jstor.org/stable/24309012

Klein, S. B., & Nichols, S. (2012). Memory and the Sense of Personal Identity. *Mind; a Quarterly Review of Psychology and Philosophy, 121*(483), 677–702. https://doi.org/10.1093/mind/fzs080

LePort, A. K. R., Mattfeld, A. T., Dickinson-Anson, H., Fallon, J. H., Stark, C. E. L., Kruggel, F., Cahill, L., & McGaugh, J. L. (2012). Behavioral and neuroanatomical investigation of Highly Superior Autobiographical Memory (HSAM). *Neurobiology of Learning and Memory, 98*(1), 78–92. https://doi.org/10.1016/j.nlm.2012.05.002

Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition, 6*(3), 312–319. https://doi.org/10.3758/BF03197461

McGregor, H. R., & Gribble, P. L. (2017). Functional connectivity between somatosensory and motor brain areas predicts individual differences in motor learning by observing. *Journal of Neurophysiology, 118*(2), 1235–1243. https://doi.org/10.1152/jn.00275.2017

Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., McAndrews, M. P., Levine, B., Black, S., Winocur, G., & Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. *Journal of Anatomy, 207*(1), 35–66. https://doi.org/10.1111/j.1469-7580.2005.00421.x

Nadel, L., Samsonovich, A., Ryan, L., & Moscovitch, M. (2000). Multiple trace theory of human memory: computational, neuroimaging, and neuropsychological results. *Hippocampus,*

10(4), 352–368. https://doi.org/10.1002/1098-1063(2000)10:4<352::AID-HIPO2>3.0.CO;2-D

Palombo, D. J., Sheldon, S., & Levine, B. (2018). Individual differences in Autobiographical Memory. *Trends in Cognitive Sciences, 22*(7), 583–597. https://doi.org/10.1016/j.tics.2018.04.007

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology. Applied, 14*(2), 101–117.

Pavlik, P. I., Jr, & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. Cognitive Science, 29(4), 559–586.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press.

Persson, J., Stening, E., Nordin, K., & Söderlund, H. (2018). Predicting episodic and spatial memory performance from hippocampal resting-state functional connectivity: Evidence for an anterior-posterior division of function. *Hippocampus, 28*(1), 53–66.

Pezzulo, G., Zorzi, M., & Corbetta, M. (2021). The secret life of predictive brains: what's spontaneous activity for? *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2021.05.007

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L., & Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron, 72*(4), 665–678.

Raichle, M. E., & Snyder, A. Z. (2007). A default mode of brain function: a brief history of an evolving idea. *NeuroImage, 37*(4), 1083–1090; discussion 1097–1099.

Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour.

Nature Reviews. *Neuroscience, 13*(10), 713–726.

Reineberg, A. E., Andrews-Hanna, J. R., Depue, B. E., Friedman, N. P., & Banich, M. T. (2015). Resting-state networks predict individual differences in common and specific aspects of executive function. *NeuroImage, 104*, 69–78.

Reineberg, A. E., & Banich, M. T. (2016). Functional connectivity at rest is sensitive to individual differences in executive function: A network analysis. *Human Brain Mapping, 37*(8), 2959–2975.

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

Sense, F., Behrens, F., Meijer, R. R., & van Rijn, H. (2016). An individual's rate of forgetting is stable over time but differs across materials. *Topics in Cognitive Science, 8*(1), 305–321.

Sense, F., Behrens, F., Meijer, R. R., & van Rijn, H. (2015). Stability of Individual Parameters in a Model of Optimal Fact Learning. *Proceedings of the 13th International Conference on Cognitive Modeling*, 136–141.

Sense, F., Maaß, S., Gluck, K., & van Rijn, H. (2019). Within-Subject Performance on a Real-Life, Complex Task and Traditional Lab Experiments: Measures of Word Learning, Raven Matrices, Tapping, and CPR. *Journal of Cognition, 2*(1), 12.

Sense, F., Meijer, R. R., & van Rijn, H. (2018). Exploration of the Rate of Forgetting as a Domain-Specific Individual Differences Measure. *Frontiers in Education, 3*, 112.

Sense, F., van der Velde, M., & van Rijn, H. (2021). Predicting university students' exam performance using a model-based adaptive fact-learning system. *Journal of Learning Analytics*, 1–15.

Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., &

Constable, R. T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, *12*(3), 506–518.

Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In M. R. Mehl (Ed.), *Handbook of research methods for studying daily life* , (pp (Vol. 676, pp. 302–320). The Guilford Press, xxvii.

Sporns, O. (2011). The human connectome: a complex network. *Annals of the New York Academy of Sciences*, *1224*, 109–125.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of Learning and Memory*, *82*(3), 171–177.

Stocco, A., Murray, N. L., Yamasaki, B. L., Renno, T. J., Nguyen, J., & Prat, C. S. (2017). Individual differences in the Simon effect are underpinned by differences in the competitive dynamics in the basal ganglia: An experimental verification and a computational model. *Cognition*, *164*, 31–45. https://doi.org/10.1016/j.cognition.2017.03.001

Stocco, A., Sibert, C., Steine-Hanson, Z., Koh, N., Laird, J. E., Lebiere, C. J., & Rosenbloom, P. (2021). Analysis of the human connectome data supports the notion of a "Common Model of Cognition" for human and human-like intelligence across domains. *NeuroImage*, *235*, 118035. https://doi.org/10.1016/j.neuroimage.2021.118035

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(26), 14792–14797. https://doi.org/10.1073/pnas.94.26.14792

Van den Broek, G. S. E., Segers, E., Van Rijn, H., Takashima, A., & Verhoeven, L. (2019). Effects of elaborate feedback during practice tests: Costs and benefits of retrieval

prompts. *Journal of Experimental Psychology: Applied*, *25*(4), 588–601. https://doi.org/10.1037/xap0000212

van der Velde, M., Sense, F., Borst, J. P., & van Rijn, H. (2020). Alleviating the Cold Start problem in adaptive learning using data-driven difficulty estimates. *Computational Brain & Behavior*. https://doi.org/10.1007/s42113-021-00101-6

Wagenmakers, E.-J., Krypotos, A.-M., Criss, A. H., & Iverson, G. (2012). On the interpretation of removable interactions: a survey of the field 33 years after Loftus. *Memory & Cognition*, *40*(2), 145–160. https://doi.org/10.3758/s13421-011-0158-0

Waller, L., Walter, H., Kruschwitz, J. D., Reuter, L., Müller, S., Erk, S., & Veer, I. M. (2017). Evaluating the replicability, specificity, and generalizability of connectome fingerprints. *NeuroImage*, *158*, 371–377.

Xu, Y., & Stocco, A. (2021). Recovering Reliable Idiographic Biological Parameters from Noisy Behavioral Data: the Case of Basal Ganglia Indices in the Probabilistic Selection Task. *Computational Brain & Behavior*, 1–17. https://doi.org/10.1007/s42113-021-00102-5

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Zhou, P., Sense, F., van Rijn, H., & Stocco, A. (2021). Reflections of idiographic long-term memory characteristics in resting-state neuroimaging data. *Cognition*, *212*, 104660. https://doi.org/10.1016/j.cognition.2021.104660