Supplementary Information

for

Multiple Traces and Altered Signal-to-Noise in Systems Consolidation:

Evidence from the 7T fMRI Natural Scenes Dataset

Thomas J. Vanasse\textsuperscript{a}, Melanie Boly\textsuperscript{a}, Emily J. Allen\textsuperscript{b,c}, Yihan Wu\textsuperscript{d}, Thomas Naselaris\textsuperscript{e}, Kendrick Kay\textsuperscript{b}, Chiara Cirelli\textsuperscript{a}, Giulio Tononi\textsuperscript{a,\Phi}

\textsuperscript{a}Center for Sleep and Consciousness, Department of Psychiatry, University of Wisconsin-Madison, Madison, WI 53719
\textsuperscript{b}Center for Magnetic Resonance Research (CMRR), Department of Radiology, University of Minnesota, Minneapolis, MN 55455
\textsuperscript{c}Department of Psychology, University of Minnesota, Minneapolis, MN 55455
\textsuperscript{d}Graduate Program in Cognitive Science, University of Minnesota, Minneapolis, MN 55455
\textsuperscript{e}Department of Neuroscience, University of Minnesota, Minneapolis, MN 55455

\textsuperscript{\Phi}Corresponding Author: 6001 Research Park Blvd, Madison, WI 53719; 608-262-7128; gtononi@wisc.edu
Beta Extraction

Betas used in this paper correspond to version “b3” as shared by Allen et al. at http://naturalscenesdataset.org and published in Nature Neuroscience. For analyses shown in Figure 2, both rep1 and rep2 trials were considered. For all following analyses we only considered rep1 trials. This was to avoid any confounding effect due to “re-consolidation” processes.

Three main steps to acquire “b3” betas correspond to a general linear model in which (1) a hemodynamic response function (HRF) is estimated for each voxel, (2) GLMdenoise was used for denoising, and (3) ridge regression was used to improve estimation of single-trial betas. While the initial public release of NSD withholds the last three NSD scan sessions from each participant, this paper includes all session data.

Here we provide some detail on the methods applied at each step. First, the best hemodynamic response function (HRF) was determined for each voxel based on an HRF library. This approach provides well-regularized HRF estimates. The second analysis component adapted GLMdenoise to a single-trial GLM framework. GLMdenoise is a technique in which data-derived nuisance regressors are identified and used to remove noise from—and, therefore, improve the accuracy of—beta estimates. The third component applied ridge regression as a method for dampening the noise inflation caused by correlated single-trial GLM predictors. To determine the optimal level of regularization for each voxel, a recently developed efficient re-parameterization of ridge regression called ‘fractional ridge regression’ was used. For each voxel, in the context of a GLM that incorporates the specific HRF chosen for that voxel, cross-validation was used to select an optimal shrinkage fraction for that voxel.

Multiple-Trace and Memory Chain Model Fitting

For MTT and Memory-Chain model fitting, we “one-hot encoded” our b3 betas-of-interest (correctly recognized rep1 trials) to the following bins:

Inter-Session Analysis Bins (Days): [0, 2, 4, 7, 10, 14, 21, 28, 35, 50, 75, 100, 125, 150, 175, 200]
Intra-Session Analysis Bins (Trials): [0, 50, 100, 150, 200, 250, 300, 350, 400]

We then applied a linear mixed-effects model using the statsmodels Python package. This model accounted for the random intercepts of subjects. Because the intercept was set to be bin 0, then linear model coefficients correspond to the relative increase in beta from that starting point (with corresponding confidence intervals). These beta coefficient, i.e. increases in activation from Trial 0 (or Day 0), were used for the memory model fits.

Memory models were fit with the curve_fit function from the scipy Python package. The python code passed to the curve_fit optimizer is provided below:

MTT Model from Nadel et al. (2000)

from scipy.integrate import quad
from numpy import np
def model_func(tau, kappa, sigma, alpha):
    return np.exp(kappa*(tau-200)) + np.exp((tau/sigma) - kappa*200)\
           *alpha*quad(integrand, np.exp(tau/sigma), np.exp(200/sigma),\
              args=(kappa,sigma))[0]
def integrand(x, a, b):
    return np.power(x,(a*b-1))/(x-1)

*Note, the time-length of 200 was switched to 400 for the within-session analysis

Memory-Chain Model from Murre et al. (2013)
def model_func(t,a1,mu2,c):
    return c*((-a1*(1-np.exp((a1)*t))**-1)/mu2 + 1)**-1

We also separately applied the lmfit package to attain the BIC values by using a least squares optimizer on the residuals.

For Figure 5, the inter-session model was fit per subject. A one-hot encoded binning procedure was again applied and an OLS model was used to extract betas for each time-point. The peak of the MTT model was identified, and the %-increase from the initial activation mean of Day 0 (intercept) was calculated (plotted in Figure 5b). To measure %-increase in within-session trials (Figure 5c), we found the percent change in activation from trials averaged 0-50 trials since recognition with that from the 325-375 trials since recognition window.

Connectivity Analyses

For the content-general connectivity analysis (Sup. Figure 2), we looked at connectivity differences among the MTL and visual system. Significant differences in rep1 connectivity (roi-to-roi Pearson correlation) between conditions of interest were assessed via a non-parametric approach by permuting outside-session and within-session recognition labels (N = 1000); a p-value was obtained by tracking where the absolute difference in correlation between conditions in the permuted data exceeded the real data. The session structure was retained in permuted data labels, i.e. the ratio between within-session vs. outside-session recognition was the same among trials for a particular session across each permutation. Only correct, rep1 betas were considered, and they were standardized per session before correlation analysis. To correct for multiple tests across all pairwise combinations, a false discovery rate BH correction (FDR p < 0.05) was applied. Among the Anterior HP, PhC, and PrC seeds there was widespread decreased connectivity (non-parametric p FDR < 0.05) with the more lateral temporal and occipital lobe aspects of the visual hierarchy (TO1,TO2,LO2; TO – temporo-occipital, LO – lateral-occipital) outside compared to within session (Sup. Figure 2). Also, the intraparietal sulci showed decreased MTL connectivity in outside-session recognition.

To perform the face-connectivity analysis we used Python’s statsmodels package. Specifically, we performed a three-way interaction test:

equation = 'roi_face ~ seed_mtl + time_since_last_rep + face + seed_x_time_since_last_rep + seed_x_face + time_since_last_rep_x_face + seed_x_time_since_last_rep_x_face'
Again, this model accounted for the random intercepts of subjects. We then investigated the significance of the seed_x_time_x_face interaction effect. We plotted violin plots of the correlation distribution between the face_roi and seed_roi using N=1000 bootstrapped (with replacement) samples separating face vs. no-face trials at both Day 0 (WS Recognition) and at the peak effect with a time cutoff of 0-20 days.

Classifier Model

We applied a logistic regression to rep1 trials using the sklearn package. The hyperparameter ‘C’ was tuned according to a nested cross-validation approach with grid search. The logistic regression was weighted to have class balance (i.e., misclassifications of the minority class were weighted higher in the loss function).

```python
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import GridSearchCV, cross_val_score, KFold

p_grid = {'C': [0.001, 0.01, 0.1, 1, 10, 100, 1000]}

lr = LogisticRegression(class_weight="balanced", solver='lbfgs')
inner_cv = KFold(n_splits=20, shuffle=True, random_state=0)
outer_cv = KFold(n_splits=40, shuffle=True, random_state=0)

clf = GridSearchCV(estimator=lr,
                   param_grid=p_grid,
                   scoring='balanced_accuracy',
                   cv=inner_cv,
                   refit=True)

nested_score_os = cross_val_score(clf,
                                   X=features,
                                   y=y,
                                   cv=outer_cv,
                                   n_jobs=n_jobs,
                                   scoring='balanced_accuracy')
```

Nested Session Variance as a Supplementary Analysis

To investigate if our results were influenced by a potential shifting baseline in BOLD activity across sessions, we performed a supplementary analysis. One way to address this potential bias is treating the session number during recognition (1-40) as a confounding nested variable (within subject). However, we see this as a problematic approach due to the sampling distribution of OS vs. WS trials. Note in Sup. Figure 7 that OS Recognition trials are skewed toward later sessions and WS Recognition trials are skewed toward early sessions. Since we observed that OS Recognition trials have significantly different BOLD activation compared to WS Trials, such
a confound analysis would likely produce an association between BOLD activation and trial number. Nevertheless, we performed this analysis.

This approach was handled with the `statsmodels` package. While all preceding analyses reported in our manuscript used Subject in a random intercepts model:

```python
import statsmodels.formula.api as smf
md = smf.mixedlm(equation, data, groups='Subject')
```

For this supplementary analysis, `recognition_session` was treated as a nested component of `subject`.

```python
vc = {'recognition_session': '0 + C(recognition_session)'}
md = smf.MixedLM.from_formula(equation, vc_formula=vc, re_formula='1', groups='Subject', data=data)
```

In this supplementary analysis we found that our reported results indeed hold for the memory model fits (see Sup. Figure 1AB), competing memory loss analysis (see Sup. Figure 1C), and the connectivity analysis (see Sup. Figure 1D).
Sup. Figure 1. Model Fits After Adjusting for Session Recognition. This analysis replicates Fig. 4 but aims to control for potential changes in ‘baseline’ signal that may occur across sessions. The session number for which each sample was recognized (1-40) was treated as a nested variable (inside ‘subject’) for the mixed-effects model. (A) shows the intra-session model fit, (B) shows inter-session model fit. (C) replicates Figure 5, also adjusting for the session of recognition. (D) replicates Figure 6, showing the significant interaction effects.
Sup. Figure 2. Significant Correlation Differences Between Outside- and Within-Session Recognition (OS – WS). A) Significant rep1 correlation differences between recognition conditions among MTL seeds and Visual System ROIs; region demonstrates significant effect at FDR corrected p<0.05 (identified by N=1000 non-parametric label shuffles retaining session-structure). See Kastner ROI labels and locations. (Significant correlation differences are colored according to directionality (Red - Within-Session Connectivity > Outside-Session Connectivity), per MTL seed, displayed on fsaverage surface. Betas of interest (correct, rep1trials) were standardized per session. B) A default threshold of p FDR < 0.05, C) corresponds to a more stringent threshold of p FDR < 0.01. Labels correspond to Kastner atlas.
Sup. Figure 3. Significant Correlation Differences Between Outside- and Within-Session Recognition (OS – WS) Among Yeo-17 Cortical Networks. Significant effects are at FDR corrected p<0.05 (N=1000 label reshuffles). Only correct rep1 trials were considered. No significant connectivity differences were found with ErC or PrC seed. Labels correspond to Yeo-17 Atlas.
Sup. Figure 4. Null Intra-Session Connectivity Effects. Seed x time (trials) x face interaction tested within a session. Contrary to the inter-session time-scale, no significant effect was found at any cutoff point within this intra-session time-scale.
Sup. Figure 5. Classifier Model Applied to Three Time-Bins. A supplementary analysis to Figure 3, but instead separating the recognition into Within-Session, Outside-Session (<20 Days), and Outside-Session (>20 Days). For F-statistics from repeated-measures ANOVA, see Sup. Table 2.
Sup. Figure 6. Within-Session vs. Outside-Session Recognition Session Distribution.
Distribution of recognition_session separated by WS Recognition rep1 trials and OS Recognition rep1 trials. OS Recognition is skewed toward later sessions, while WS Recognition is skewed toward early sessions.
| Subject ID | Birth Year | Age | Sex | Race                   | Height (inches) | Weight (pounds) | Right-handed? |
|-----------|------------|-----|-----|------------------------|----------------|-----------------|---------------|
| Subj01    | 1989       | 30  | M   | White, Not Hispanic    | 72             | 149             | yes           |
| subj02    | 1991       | 28  | F   | White, Not Hispanic    | 66             | 140             | yes           |
| subj03    | 1990       | 29  | F   | White, Not Hispanic    | 65             | 135             | yes           |
| subj04    | 1992       | 27  | F   | Asian, Not Hispanic    | 62             | 121             | yes           |
| subj05    | 1987       | 32  | F   | Asian, Not Hispanic    | 62             | 138             | yes           |
| subj06    | 1996       | 23  | M   | White, Not Hispanic    | 68             | 115             | yes           |
| subj07    | 1995       | 24  | F   | Asian, Not Hispanic    | 63             | 105             | yes           |
| subj08    | 2000       | 19  | F   | White, Not Hispanic    | 64             | 145             | yes           |

**Supplementary Table 1. Subject Demographics**
**Sup. Table 2. Statistics of Classifier Model Applied to Three Time-Bins.** F-statistics of classifier analysis at three time points: WS Recognition, OS Recognition (Days < 20), OS Recognition (>20 days). Only the main effect of Features was significant at P<0.001.

|                  | F-Value | P     |
|------------------|---------|-------|
| Session          | 1.97    | 0.175 |
| Features         | 85.23   | <0.0001 |
| Session x Features | 0.5319 | 0.7133 |
**Sup. Table 3. Trials in Face Connectivity Analysis.** (Top) The number of trials considered in FACE vs. NO FACE categories for the connectivity analysis at separate “cutoff” days. (Bottom) The percent increase (from Day 0) in both categories.

| Day Cutoff | 0     | 10    | 20    | 30    |
|------------|-------|-------|-------|-------|
| FACE Trials| 5864  | 7018  | 7400  | 7715  |
| NO FACE Trials| 12247 | 14591 | 15312 | 15931 |

% Increase from Day 0

|        | FACE |        |        |       |
|--------|------|--------|--------|-------|
| 0%     | 19.68% | 26.19% | 31.57% | 0%    |
| 0%     | 19.14% | 25.03% | 30.08% | 0%    |
Supporting Information References

1. Allen, E. J. et al. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nat Neurosci* 1–11 (2021) doi:10.1038/s41593-021-00962-x.

2. Kay, K., Rokem, A., Winawer, J., Dougherty, R. & Wandell, B. GLMdenoise: a fast, automated technique for denoising task-based fMRI data. *Frontiers in Neuroscience* 7, (2013).

3. Rokem, A. & Kay, K. Fractional ridge regression: a fast, interpretable reparameterization of ridge regression. *Gigascience* 9, giaa133 (2020).

4. Seabold, S. et al. statsmodels/statsmodels: Version 0.8.0 Release. *Zenodo* (2017) doi:10.5281/zenodo.275519.

5. Virtanen, P. et al. scipy/scipy: SciPy 1.2.0. *Zenodo* (2018) doi:https://doi.org/10.5281/zenodo.2377507.

6. Newville, M., Stensitzki, T., Allen, D. B. & Ingargiola, A. *LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python*. (2021).

7. Grisel, O. et al. scikit-learn/scikit-learn: scikit-learn 1.0.1. *Zenodo* (2021) doi:https://doi.org/10.5281/zenodo.5596244.

8. Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830 (2011).