Supplementary Information for

Identifying Causal Subsequent Memory Effects

David J. Halpern, Shannon Tubridy, Lila Davachi and Todd M. Gureckis

David Halpern.
E-mail: david.halpern@nyu.edu

This PDF file includes:

- Supplementary text
- Figs. S1 to S15 (not allowed for Brief Reports)
- SI References
Supporting Information Text

Causal Models

This section lays out a causal graphical model approach to understand how causal encoding activity can be identified (1).

Concreteness Model. We first imagine, as in the main text, the situation if the only variable that causally impacted memory (through neural activity) was concreteness. In this model, shown graphically in figure S1, we assume that the concreteness of a stimulus drives neural activity and some of that activity, labeled Causal Encoding Activity, causes the probability of later memory to increase or decrease. This is downstream from concreteness, and can presumably be thought of as something like "perceived concreteness." Both Causal Encoding Activity and Non-causal Encoding Activity likely contribute to any particular fMRI signal we are interested in. The key question, then, is how to identify fMRI signals that measure Causal Encoding Activity, or in other words, signals where the red arrow in figure S1 exists. One useful thing we can derive from graphs, using the logic laid out by (1), are the implied (conditional) independences. This logic provides easy algorithms for checking dependencies in graphs and they are implemented in packages like dagitty in R (2). In the graph without the red arrow, when we condition on Concreteness, the fMRI signal will be independent of Memory. This suggests that we can distinguish fMRI signals that measure Causal Encoding Activity from those that do not by testing whether fMRI signals that are not independent from Memory after conditioning on Concreteness.

Endogenous-Exogenous Model. Following (3), we divide up the factors affecting memory into endogenous and exogenous factors. Exogenous factors represent a group of variables that are in principle under experimenter control because they involve features of the item presented or features of the way in which items are presented. Many such features have been extensively documented in the psychological literature as having effects on memory performance. Since these are observable, through the method we describe in the paper, we can simply replace Concreteness in Fig. S1 with the effects of all Exogenous variables as in Fig. S2. In this case, because we can block the path from the non-causal neural activity to memory, we can use our approach to identify causal encoding activity.

In contrast, endogenous features represent ongoing latent cognitive or neural processes that are known to affect memory such as fluctuations in attention or fatigue, or their neural correlates. These processes are necessarily latent and it is difficult to obtain precise measures of them from behavior alone. Figure S3 shows a causal graph in which both sets of variables influence memory. The endogenous variables (which may include causal activity from previous trials) are represented as a dotted circle because we do not measure them. What becomes clear from this graph is that it is now complicated to separate out the causal activity from the non-causal. This is because if some endogenous neural activity prior to the trial caused both causal and non-causal neural activity to change, without controlling for the endogenous factors, both sources of activity would appear to be correlated with memory.

To clarify exactly which non-causal neural activity will appear correlated with memory, we expand our model further as in Fig. S4. Here, we can separate our endogenous variables into causal and non-causal. Causal endogenous variables represent variables that cause encoding activity (and indirectly, memory performance) while non-causal endogenous variables do not. Only signals that reflect causal neural activity or activity downstream from causal neural activity will appear correlated with memory after controlling for the exogenous variables. Therefore, we call these indicators of causal encoding activity. As mentioned in the main text, these signals are still valuable to discover as they provide measurements of causal endogenous variables, which may be useful in constructing theories. In addition, because this is a subset of all non-causal encoding activity, this approach restricts the set of signals of interest for future studies.

Subsequent memory GLMs

To complement the main text, this section includes contrast maps showing the results of subsequent memory analyses obtained from general linear models (GLMs) of the fMRI data. The fMRI data were analyzed using a two-stage approach. We first estimated the within-participant effects of interest and then computed the group-level across participant stability of the effect for each voxel (4, 5).

First level within-participant GLMs. Individual participants’ preprocessed BOLD data were analyzed using a general linear model (GLM) as implemented in SPM12 (4) *. The BOLD activation for each of the five study runs was modeled in a single regression with the design matrix composed of: a predictor with a binary coding corresponding to the onset of study trials for words that were given correct responses on the recall test (remembered); a predictor with a binary coding of the onset of study trials for words that were given incorrect responses on the recall test (forgotten); a set of fourteen potential nuisance and confound signals as estimated during preprocessing (the six motion and realignment estimates obtained during preprocessing and an additional eight predictors carrying signal from white matter and CSF (6)); and five variables coding for the time points spanning each of the five separately recorded study runs. The predictors of interest (remembered, forgotten) were constructed as four-second boxcars covering the timepoints of their occurrence for each participant. Prior to model estimation these vectors were convolved with the canonical SPM double gamma hemodynamic response function (HRF) along with its temporal and dispersion derivatives (7) to account for the timing of the BOLD signal relative to stimulus events. The nuisance confounds, which were estimated from the data BOLD data, were not convolved with the HRF. The GLM was then fit using ordinary

* [https://www.fil.ion.ucl.ac.uk/spm/software/spm12/](https://www.fil.ion.ucl.ac.uk/spm/software/spm12/)
least squares (OLS) resulting in regression coefficients for each predictor of interest for each voxel summarizing the relationship of these predictors to the voxel time series during the course of the study session.

**Second level group analysis.** To assess the group-level stability of the effects estimated in the first level within-participant models, we calculated within participant linear contrasts of the regression coefficients for remembered compared to forgotten trials in a brain-wide contrast volume for each participant. These contrast images were then carried forward to the second level where a one-sample t-test across participants assessed the significance of the contrast at the group level for each voxel. This procedure was repeated for each contrast of interest. For supplemental visualization the maps were thresholded at voxel-wise p<.001 and a 5 voxel cluster threshold. Contrast maps were generated collapsing across all five study events within remembered and forgotten bins as well as for remembered vs forgotten contrasts within each of the five study events.

**Regions showing subsequent memory effects.** The group GLM-derived contrasts of remembered versus forgotten trials averaged across all five study trials yielded a number of cortical regions showing greater BOLD signal on subsequently remembered compared to subsequently forgotten trials (Figure S5). The identified voxels were in bilateral inferior parietal cortex along the inferior parietal sulcus; left pre-motor cortex, frontal cortex, and anterior insula (8); bilateral fusiform gyrus; bilateral inferior temporal gyrus; left occipito-temporal cortex overlapping with the so-called Visual Word Form Area (9).

Examining the remembered versus forgotten contrasts for only the first study trial for each word pair (Figure S6), we observed a substantial overlap with the maps from all study trials along with a few differences. The right parietal and inferior temporal regions in the all study trials contrast were not present in the first study trial contrast. Conversely, the extent of left prefrontal activation was broader in the first trial compared to all trials. In addition, there were voxels in the left caudate, around the amygdala-hippocampus border and superior amygdala, and along the medial portion of the left superior frontal gyrus that were not identified in the all study trials contrast.

The remembered versus forgotten contrasts carried out separately for study trials 2 through 5 (Figures S7, S8, S9, and S10) showed a substantial reduction in above-threshold voxels, with a left insula and right parietal region remaining present in study trials 2 and 3 respectively before dropping out on the final two study attempts.

**Individual Study Trial SMEs and ICEA SMEs**

In addition to the maps generated using the mass univariate SME contrast, we can also run the same models as in the main text using only data from a single study trial or study trial-pair. Figure S14 shows the standard SME results using individual trials. We find many effects, including activity on the first four trials in all Schaefer regions predict success memory at standard alpha levels but after adjusting for the false discovery rate, only two results are significant: activity on the third trial in the Targeted regions and Global Pattern Similarity in the hippocampus between the 1st and 3rd trial. These results, compared to the results in the main text, likely indicate the need to integrate across the entire study period when making subsequent memory predictions. Figure S15 shows the results for the ICEA SME. Here again, several features are significant but none after adjusting for the false discovery rate, suggesting a lack of reliable evidence for an ICEA SME. Interestingly, like in the main results, to the extent that a significant result here might indicate an interesting feature for future studies to examine, there does not seem to be much of a relationship between the ICEA SME and the standard SME results, suggesting that these two analyses pick up on very different signals.

**Comparison of Online and Lab Memory Tests**

As mentioned in the Methods section of the main text, 13 of the 57 subjects in our study completed their final memory test online. While we analyzed these data combined in the main text, we here conduct analyses to see if there is evidence that Online subjects performed our task differently than Lab subjects, possibly contributing to our inability to detect causal subsequent memory effects. Figure S11a shows the distributions of subject performance in both groups and Figure S11b compares estimates of memorability for each item among Online and Lab subjects. In order to test for differences statistically, we fit the following Bayesian logistic mixed effects model: $P(r_{ws} = 1|s, w) = \logit^{-1}(\alpha + (\beta + \gamma_w)l_s + \theta_l + \eta_w)$ with $\theta_l \sim N(0, \sigma_\theta)$ and $\eta_w \sim N(0, \sigma_\eta)$ and $\gamma_w \sim N(0, \sigma_\gamma)$. Here, w indexes Lithuanian-English word pairs, s indexes subjects, $l_s$ is the test location for each subject and $r_{ws}$ is the response (correct or incorrect recall) of each subject to each word pair. $\beta$ is the overall difference in performance between subjects who tested Online vs. in Lab and $\gamma_w > 0$ indicates evidence that word memorability differed across the two test locations. This model allows for differential item functioning between the two groups (3) we fit this model using the rstanarm package (10), using the default prior settings of $\alpha \sim N(0,2.5)$, $\beta \sim N(0,6)$, $\sigma_\theta$, $\sigma_\gamma$, and $\sigma_\eta \sim \text{Gamma}(1, 1)$. We ran the dynamic Hamiltonian Monte Carlo algorithm (11) for 2500 iterations. This resulted in no divergent transitions (11), split potential scale reduction factor $(\hat{R}) < 1.01$ and rank-normalized effective sample sizes of > 400 for all parameters (12), suggesting that the sampled parameters were sufficient for inference. Figure S12 shows parameter estimates and 95% posterior intervals for key parameters. We find that the key $\beta$ parameter is very likely to be positive although the low end of the posterior interval is near 0. Looking at the plot in Fig. S11a suggests that this is likely due to a lack of low performers who got fewer than 10% of word pairs correct in the Online group. While we cannot know for sure from our existing data, these differences may have arisen due to the Online data being collected later in time (and potentially at a different time during the semester) or because completing the test required an additional level of technological savvy. Alternately, with a sample of only 13 in one of the groups, this difference could be simply a rare chance event. Crucially, however, we see that $\sigma_\eta$ has a high probability of being either equal to or very close to 0. This suggests that the word-pair
memorabilities differed very minimally between groups, which is consistent with the high degree of correlation in Fig. S11b. We interpret this result as suggesting that Online subjects may have had slightly higher memory ability than Lab subjects but did not use meaningfully different memory strategies and therefore there was no differential item functioning. Because fitting a frequentist mixed effects model (using lmer) with differential item functioning resulted in a singular fit (and a maximum likelihood estimate of 0 variance), we thus concluded that removing these random effects resulted in an acceptable model.

Power Analysis

As our primary result in this paper is ultimately a failure to detect a causal subsequent memory effect, a key question is whether we had sufficient power to expect to detect it. While our study had many more subjects than nearly all other subsequent memory studies we are aware of, we had relatively limited trials, compared to other studies using recognition memory tasks after shorter delays (e.g. (13)), due to our desire to have adequate cued recall performance after 72 hours. To simplify and make our analysis more similar to analyses in the literature, we conduct a post-hoc power analysis using effect sizes from a single neural signal that had a reasonably strong standard subsequent memory effect in our data, the average activity in the left ventromedial prefrontal cortex. To estimate this, we first fit the following mixed effects model in lmer (14):

$$\text{activity}_{LVMPFC} \sim \text{recalled} + (\text{recalled} | \text{subject})$$.

This gave us a subsequent memory effect estimate (the $\beta$ on recalled) using a standard approach. We then derived the variation in this effect from a model that also allowed the effect to vary by item (where the effect of recalled on activity$_{LVMPFC}$ was no longer significant) activity$_{LVMPFC} \sim \text{recalled} + (\text{recalled} | \text{subject}) + (\text{recalled} | \text{item})$.

This gave us variance terms for both subjects and items. Finally, we fit a model of recall using a logistic mixed effects model recalled $\sim (1 | \text{subject}) + (1 | \text{item})$. Using the parameters derived from this model, we simulated the recalled data vector for arbitrary numbers of subjects and items. We then used the parameters from the two subsequent memory models ($\beta$ from the standard model and variance terms from the ICAE model) to simulate the activity$_{LVMPFC}$ data vector. We chose numbers of subjects and items based on the parameters of experiments in the subsequent memory studies we cited in the paper. All except (15) used fewer than 25 subjects and while some used 400 items (e.g. early studies such as (13) and (16)) most used much fewer (in particular, while having a high number of subjects, (15) only used 40 items). The results of this power analysis suggest that our study with 57 subjects and 45 items had greater power (68%) to detect a subsequent memory effect than a hypothetical study with similar effect sizes that used the design of other papers we cited as evidence of subsequent memory effects (Fig. S13). However, even for a subsequent memory effect, we were likely under-powered relative to standard of 80%.

Overall, this is consistent with the idea that current SME studies are not precise enough to identify a causal SME. The result that our study had high power relative to a hypothetical study with greater numbers of items is intuitive because the effect of increasing the number of samples is nonlinear and concave. Under some simplifying assumptions, such as that there is equal variance in the SME across subjects as across items, going from a study with 13 subjects to 57 subjects is a much greater increase in statistical precision than going from 45 items to 480 items (as in (13)). This can be seen because $\frac{1}{13} - \frac{1}{57} > \frac{1}{45}$, meaning that no increase in the number of items could be as big as the difference between 13 and 57.

Comparison to Previous Work

While the most common designs in subsequent memory analyses focuses on neural signals obtained during stimulus presentation, a number of studies have used alternative approaches to understand how activity during an encoding period more broadly might relate to subsequent memory performance. While not explicit about the goal of identifying causal activity, many of these approaches do not suffer from as many confounders as the classic approach because they do not focus on stimulus-evoked activity.

One set of studies has investigated longer-lasting mnemonic states by averaging signals over several temporally contiguous encoding events (e.g. (3, 17)). While in principle these results should be less affected by item-level effects, small batches of items can still vary strongly in their overall memorability. Thus, while there is likely less variance to explain with item/task-related neural effects, it is still possible for an SME in these approaches to be driven by confounds. In a related approach, Otten et al. (18) studied the effect of longer periods of activity after regressing out the item-level subsequent memory effect. However, high performance periods are likely due, at least in part, to interactions between items (i.e. context-effects) which cannot be adjusted for using this approach.

Another set of studies has investigated subsequent memory effects in activity prior to a stimulus being presented (e.g. (19, 20)). As pre-stimulus activity cannot be driven by semantic/perceptual features of the stimulus, these effects are therefore more likely to represent ICEA. However, pre-stimulus activity may still be driven by task features such as serial position, which also is a strong predictor of memory. Additionally, because of the known contiguity effect, which affects not only free recall but also recognition (21) and cued recall (22), the memorability of the previous item (which could be captured by the pre-stimulus activity) will predict the subsequent memory of the current stimulus. Overall, these types of designs are less likely to be affected by the issues discussed in the present work but they have not completely eliminated the possibility that results are driven by confounding variables. Therefore, we suggest future work in this vein investigating ICEA would benefit from an approach that allows for statistical adjustment such as the one used in the present work.

As mentioned in the main text, Weidemann & Kahana (3) have recently conducted a similar study to ours, statistically adjusting item-level EEG subsequent memory effects for stimulus memorability and serial position effects. While the magnitude of their subsequent memory effects are reduced after adjustment, they still find evidence of a predictive neural signal beyond these potential confounders. Here we speculate on why our findings may differ from theirs, pointing out in which ways our
studies diverge. First and most obvious is the recording technology. Oscillatory signals in scalp EEG could be capturing slightly different neural activity that is closer to the kind of causal activity we are interested in. Second, Weidemann & Kahana model their data at an individual level and only aggregate the accuracies at the subject level. This means that the learned model may have ultimately been very different for each subject. While this may make the model more sensitive, their significant results do not imply that there is any consistent population-level effect, which would be required for our inter-subject modeling approach to be successful. Third, Weidemann & Kahana’s statistical approach does not allow for accounting for the context level effects, i.e. potential interactions between items, unlike ours here. Thus, there may be some residual unaccounted for confounding in their analysis for identifying ICEA.

Finally, the subsequent memory literature is extensive and has many links to work with lesion patients and animals as well as purely anatomical studies. In particular, a few studies have moved beyond documenting the existence of SMEs in various regions to testing theory-driven hypotheses about the ways in which neural structures contribute to memory encoding, such as dissociations in the direction of the effect across regions. Despite using a standard subsequent memory approach, this work may seem to identify a priori more plausible causal activity. Here we discuss two papers of this type and the possibility of accounting for their findings with confounding variables as in the present work.

In Davachi et al. (16), subjects were told to encode a word while either imagining it in a specific place or simply reading it and were later given item recognition as well as source recollection tests for the task used during encoding. Based on a combination of anatomy, lesion and animal work, Davachi et al. reasoned that hippocampus and parahippocampal cortex would be related to associative memory while perirhinal cortex would be related to recognition or item memory. Looking at only the trials in the imagine condition, they found that hippocampal and posterior parahippocampal activity indeed varied only with source recollection performance, while perirhinal cortex varied only with item recognition performance, as predicted. In another study, LaRocque et al. (23), based on a number of computational models as well as neuropsychological data, predicted that hippocampus and the surrounding MTL would differ in their degree of pattern separation vs. pattern completion during encoding of items. Specifically, they predicted that neural similarity across items would relate in opposite ways to subsequent memory in hippocampus compared to perirhinal and parahippocampal cortex. This was indeed born out by the data and both of these works can be considered a success for theory-driven cognitive neuroscience.

Nonetheless, we can still ask whether there are plausible alternative accounts in terms of the types of variables discussed in the present work. Broadly, item memorability, serial position effects, and other memory-related variables are not single constructs but rather can depend on the way in which memory is queried (e.g., recognition vs. recall or associative vs. item), the manner in which it was studied, or other such factors meaning that memorability effects can vary across conditions. While there are several possible memorability-based explanations for the preceding findings, we will present just an example for each. In Davachi et al. (16), where the subsequent memory analyses were focused on those trials where participants engaged in imagery, it is possible that an item being easily image-able or concrete led to a higher probability of remembering that it was imagined and also altered certain specific neural responses but not others (24). In LaRocque et al. (23), it is possible that memorable items have stronger category representations in some regions (PRc, PHc) and more distinct representation in others (hippocampus). Additionally, more memorable items might have more stable neural representations which could also drive the results in these types of pattern analyses. We want to emphasize that we do not believe these explanations are more plausible than the theory-driven explanations that led to the genesis of these papers. Rather, without the proper design and analysis, the data cannot really provide evidence for one explanation or the other. In this paper, we try to lay out a set of criteria for doing so, which we hope will provide a better foundation in the future.
Fig. S1. A directed acyclic graph representing a causal model for memory if abstractness was the only variable that influenced memory. Solid circles indicate observed variables while dotted circles indicate unobserved.
Fig. S2. A directed acyclic graph representing a causal for memory if exogenous variables were the only variables that influenced memory. Solid circles indicate observed variables while dotted circles indicate unobserved.
Fig. S3. A graph including all endogenous and exogenous variables that affect memory. Solid circles indicate observed variables while dotted circles indicate unobserved.
Fig. S4. A graph showing the distinction between causal encoding activity and indicators of causal encoding activity. Using our approach, we cannot distinguish between these two types of variables, due to the fact that the endogenous variables are uncontrolled. Solid circles indicate observed variables while dotted circles indicate unobserved.
Fig. S5. Contrast maps (t-statistics) for Remembered > Forgotten (warm) and Forgotten > Remembered (cool) contrasts from group level GLMs collapsing across five study opportunities (All study trials). The maps are thresholded at voxel-wise p < .001, uncorrected, with 5 voxel minimum cluster threshold. The MNI z coordinates are shown.
Study trial 1

Fig. S6. Contrast maps (t-statistics) for Remembered > Forgotten (warm) and Forgotten > Remembered (cool) contrasts on the first study repetition. The maps are thresholded at voxel-wise p < .001, uncorrected, with 5 voxel minimum cluster threshold. The MNI z coordinates are shown.
Study trial 2

Fig. S7. Contrast maps (t-statistics) for Remembered > Forgotten (warm) and Forgotten > Remembered (cool) contrasts on the second study repetition. The maps are thresholded at voxel-wise p < .001, uncorrected, with 5 voxel minimum cluster threshold. The MNI z coordinates are shown.
Study trial 3

Fig. S8. Contrast maps (t-statistics) for Remembered > Forgotten (warm) and Forgotten > Remembered (cool) contrasts on the third study repetition. The maps are thresholded at voxel-wise p < .001, uncorrected, with 5 voxel minimum cluster threshold. The MNI z coordinates are shown.
Study trial 4

Fig. S9. Contrast maps (t-statistics) for Remembered > Forgotten (warm) and Forgotten > Remembered (cool) contrasts on the fourth study repetition. The maps are thresholded at voxel-wise p < .001, uncorrected, with 5 voxel minimum cluster threshold. The MNI z coordinates are shown.
Study trial 5

Fig. S10. Contrast maps (t-statistics) for Remembered > Forgotten (warm) and Forgotten > Remembered (cool) contrasts on the fifth study repetition. The maps are thresholded at voxel-wise p < .001, uncorrected, with 5 voxel minimum cluster threshold. The MNI z coordinates are shown.
Fig. S11. Behavioral data split by subjects who completed the final memory test in Lab or Online. a shows the distribution of subject performance in each test location. b shows the estimates of memorability for each word pair among Lab subjects (x axis) and Online subjects. Estimates and confidence intervals are computed using the Wilson method (25, 26), due to the small sample sizes and raw proportions close to 1.
Fig. S12. Posterior distributions for key parameters in logistic linear mixed effects model. $\beta$ corresponds to the difference between Online and Lab subjects in terms of overall performance. $\sigma_\gamma$ corresponds to the variance in this difference across word-pairs. Thicker bars indicate the central 50% posterior region and thin bars indicate 95%.
Fig. S13. Power in our dataset for several experimental designs (combinations of numbers of subjects and items.)
Fig. S14. Classifier performance based on the Standard Subsequent Memory Model using individual study blocks or study block pairs. Each combination of ROIs, features and time-point treatment is plotted separately with definitions of terms found in Methods. IPS = Item Pattern Similarity, GPS = Global Pattern Similarity, ISPC = Intersubject Pattern Correlation. Black line indicates chance AUC (.5) and statistical tests are compared with this baseline. * = \( p < .05 \) based on a permutation t-test, ** = \( q < .05 \) after false discovery rate (FDR, 27). Error bars reflect the unadjusted standard error of the mean.
Fig. S15. Classifier performance based on the *ICEA Subsequent Memory Model* using individual study blocks or study block pairs. Each combination of ROIs, features and time-point treatment is plotted separately with definitions of terms found in Methods. IPS = Item Pattern Similarity, GPS = Global Pattern Similarity, ISPC = Intersubject Pattern Correlation, JOL = Judgment of Learning, IRT = Item Response Theory model. JOL and IRT are plotted in each column for comparison although they do not differ across columns. * = p < .05 based on a permutation t-test, ** = p < .05 after an FDR (27) adjustment. Error bars reflect the unadjusted standard error of the mean.
References

1. J Pearl, Causal diagrams for empirical research. *Biometrika* **82**, 669–688 (1995).
2. J Textor, B van der Zander, MS Gilthorpe, M Liškiewicz, GT Ellison, Robust causal inference using directed acyclic graphs: the R package ‘dagitty’. *Int. J. Epidemiol.* **45**, dyw341 (2017).
3. CT Weidemann, MJ Kahana, Neural measures of subsequent memory reflect endogenous variability in cognitive function. *J. Exp. Psychol. Learn. Mem. Cogn.* **47** (2020).
4. KJ Friston, et al., Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* **2**, 189–210 (1994).
5. JA Mumford, T Nichols, Simple group fMRI modeling and inference. *NeuroImage* **47**, 1469–1475 (2009).
6. Y Behzadi, K Restom, J Liu, TT Liu, A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage* **37**, 90–101 (2007).
7. MA Lindquist, J Meng Loh, LY Atlas, TD wager, Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling. *NeuroImage* **45**, S187–S198 (2009).
8. H Kim, Neural activity that predicts subsequent memory and forgetting: A meta-analysis of 74 fMRI studies. *NeuroImage* **54**, 2446–2461 (2011).
9. L Chen, et al., The visual word form area (VWFA) is part of both language and attention circuitry. *Nat. Commun.* **10**, 5601 (2019).
10. B Goodrich, J Gabry, J Alì, S Brilleman, rstanarm: Bayesian applied regression modeling via Stan. (2020).
11. M Betancourt, A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv:1701.02434 [stat] 1* (2018) arXiv: 1701.02434.
12. A Vehtari, A Gelman, D Simpson, B Carpenter, PC Bürkner, Rank-Normalization, Folding, and Localization: An Improved $\hat{R}$ for Assessing Convergence of MCMC. *Bayesian Analysis* **16** (2020).
13. AD Wagner, Building Memories: Remembering and Forgetting of Verbal Experiences as Predicted by Brain Activity. *Science* **281**, 1188–1191 (1998).
14. D Bates, M Mächler, B Bolker, S Walker, Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **67** (2015).
15. M Aly, NB Turk-Browne, Attention promotes episodic encoding by stabilizing hippocampal representations. *Proc. Natl. Acad. Sci.* **113**, E420–E429 (2016).
16. L Davachi, JP Mitchell, AD Wagner, Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proc. Natl. Acad. Sci.* **100**, 2157–2162 (2003).
17. G Fernandez, JB Brewer, Z Zhao, GH Glover, JD Gabrieli, Level of sustained entorhinal activity at study correlates with subsequent cued-recall performance: A functional magnetic resonance imaging study with high acquisition rate. *Hippocampus* **9**, 35–44 (1999).
18. LJ Otten, RNA Henson, MD Rugg, State-related and item-related neural correlates of successful memory encoding. *Nat. Neurosci.* **5**, 1339–1344 (2002).
19. LJ Otten, AH Quayle, S Akram, TA Ditewig, MD Rugg, Brain activity before an event predicts later recollection. *Nat. Neurosci.* **9**, 489–491 (2006).
20. ZJ Urgolites, et al., Spiking activity in the human hippocampus prior to encoding predicts subsequent memory. *Proc. Natl. Acad. Sci.* **117**, 202001338 (2020).
21. G Schwartz, MW Howard, B Jing, MJ Kahana, Shadows of the Past: Temporal Retrieval Effects in Recognition Memory. *Psychol. Sci.* **16**, 898–904 (2005).
22. OC Davis, AS Geller, DS Rizzuto, MJ Kahana, Temporal associative processes revealed by intrusions in paired-associate recall. *Psychon. Bull. & Rev.* **15**, 64–69 (2008).
23. KF LaRocque, et al., Global Similarity and Pattern Separation in the Human Medial Temporal Lobe Predict Subsequent Memory. *J. Neurosci.* **33**, 5466–5474 (2013).
24. D Hassabis, D Kumaran, EA Maguire, Using Imagination to Understand the Neural Basis of Episodic Memory. *J. Neurosci.* **27**, 14365–14374 (2007).
25. EB Wilson, Probable Inference, the Law of Succession, and Statistical Inference. *J. Am. Stat. Assoc.* **22**, 209–212 (1927).
26. A Agresti, BA Coull, Approximate is Better than “Exact” for Interval Estimation of Binomial Proportions. *The Am. Stat.* **52**, 119–126 (1998).
27. Y Benjamini, Y Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Royal Stat. Soc. Ser. B (Methodological)* **57**, 289–300 (1995).