Published in final edited form as:
Neuropsychologia. 2012 November ; 50(13): 3015–3026. doi:10.1016/j.neuropsychologia.2012.07.010.

Mechanisms supporting superior source memory for familiar items: A multi-voxel pattern analysis study

Jordan Poppenk[a,*] and **Kenneth A. Norman**[a,b]

[a]Princeton Neuroscience Institute, Green Hall, Princeton University, Princeton, NJ, USA, 08540

[b]Department of Psychology, Green Hall, Princeton University, Princeton, NJ, USA, 08540

Abstract

Recent cognitive research has revealed better source memory performance for familiar relative to novel stimuli. Here we consider two possible explanations for this finding. The source memory advantage for familiar stimuli could arise because stimulus novelty induces attention to stimulus features at the expense of contextual processing, resulting in diminished overall levels of contextual processing at study for novel (vs. familiar) stimuli. Another possibility is that stimulus information retrieved from long-term memory (LTM) provides scaffolding that facilitates the formation of item-context associations. If contextual features are indeed more effectively bound to familiar (vs. novel) items, the relationship between contextual processing at study and subsequent source memory should be stronger for familiar items. We tested these possibilities by applying multi-voxel pattern analysis (MVPA) to a recently collected functional magnetic resonance imaging (fMRI) dataset, with the goal of measuring contextual processing at study and relating it to subsequent source memory performance. Participants were scanned with fMRI while viewing novel proverbs, repeated proverbs (previously novel proverbs that were shown in a pre-study phase), and previously known proverbs in the context of one of two experimental tasks. After scanning was complete, we evaluated participants' source memory for the task associated with each proverb. Drawing upon fMRI data from the study phase, we trained a classifier to detect on-task processing (i.e., how strongly was the correct task set activated). On-task processing was greater for previously known than novel proverbs and similar for repeated and novel proverbs. However, both within- and across participants, the relationship between on-task processing and subsequent source memory was stronger for repeated than novel proverbs and similar for previously known and novel proverbs. Finally, focusing on the repeated condition, we found that higher levels of hippocampal activity during the pre-study phase, which we used as an index of episodic encoding, led to a stronger relationship between on-task processing at study and subsequent memory. Together, these findings suggest different mechanisms may be primarily responsible for superior source memory for repeated and previously known stimuli. Specifically, they suggest that prior stimulus knowledge enhances memory by boosting the overall level of contextual processing, whereas stimulus repetition enhances the probability that contextual features will be successfully bound to item features. Several possible theoretical explanations for this pattern are discussed.

[*]Correspondence and requests for materials should be addressed to J. Poppenk, Phone: +1 609 356 8708, jpoppenk@princeton.edu.

The authors declare that they have no competing financial interests.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

novelty; repetition; prior knowledge; episodic memory; multi-voxel pattern analysis (MVPA); fMRI

1. Introduction

Studies revealing recognition memory advantages for novel stimuli over repeated stimuli (i.e., stimuli that were presented in an earlier, “pre-study” phase of the experiment; see, e.g., Tulving & Kroll, 1995) have been influential in cognitive neuroscience, inspiring several theories about neural mechanisms that could induce superior encoding of novel information (e.g., Lisman & Grace, 2005; Tulving, Markowitsch, Craik, Habib, & Houle, 1996). Recent cognitive data, however, indicate that prior observations of superior memory for novel over repeated stimuli occurred because of factors relating to retrieval, rather than encoding: Specifically, these data indicate that classic findings of worse memory for repeated items arise from source confusion (mistaking stimuli learned elsewhere for stimuli from the study phase) rather than poor learning of study phase details *per se* (Poppenk, Köhler & Moscovitch, 2010a). The question of how novelty affects encoding of study-phase details can be more directly addressed using source memory tests that probe memory for unique contextual details from the study phase (e.g., the encoding task that was performed on the item at study). By focusing the test on details that pertain only to the study phase (as opposed to other events involving the queried item), the potential for source confusion at retrieval is greatly reduced, making it possible to examine – in a relatively unconfounded fashion – effects of novelty versus familiarity on encoding. Numerous studies using this kind of source memory test have found better source memory for repeated vs. novel proverbs, scenes, faces and words (Lee, Jung, & Yi, 2012; Poppenk et al., 2010a; Poppenk, McIntosh, Craik & Moscovitch, 2010b; but see Kim et al., 2012). Furthermore, Poppenk et al. (2010a) found a similar source memory bonus when proverbs known to participants from prior life experience were compared to novel items. These findings call for identification of a mechanism that can explain this source memory bonus for familiar (i.e., repeated or previously known) stimuli.

Here, we consider two (non-exclusive) families of explanations for the familiarity bonus that has been observed in these studies, and then we evaluate these explanations using multivariate pattern analysis applied to an fMRI dataset from Poppenk & Moscovitch (2011). Intuitively, there are two prerequisites that must be met at encoding in order to support good source memory: First, participants must process the relevant contextual (source) features. Second, to retrieve contextual features at test, these contextual features must be successfully bound to the representation of the item. As discussed below, stimulus familiarity can affect both of these steps (initial processing of contextual features, and successful binding of these features to the item representation).

One possible explanation for the familiarity bonus is that stimulus novelty at encoding diverts attentional resources towards processing of the (novel) stimulus and away from processing of contextual details (e.g., encoding task or spatial position), resulting in diminished overall levels of contextual processing at study for novel (vs. familiar) stimuli. This idea encompasses a family of mechanisms that we will describe collectively as the *attention* account. The key shared feature of all of these mechanisms is the prediction that less processing of contextual features should occur for novel vs. familiar stimuli. Attentional diversion could have a negative impact on source memory in a number of ways: Divided attention at encoding is known to have negative effects on memory (Craik, Govoni, Naveh-Benjamin, & Anderson, 1996; especially associative memory, Castel & Craik, 2003) and explicitly directing attention towards item features during memory encoding reduces subsequent source memory (Dulas, 2011). Such effects are consistent with the encoding specificity principle (Tulving & Thomson, 1973), which posits that the type of information processing that takes place at encoding will determine the type of information available in memory at test. Along these lines, it has been proposed that the hippocampus captures only the contents of consciously apprehended information in forming episodic memories (i.e., a form of memory that incorporates vivid reinstatement of the contextual details of an event; Moscovitch, 2008). Accordingly, to the extent that attention is diverted away from contextual information and towards item information, contextual details should be lost. There is at least some evidence in the literature suggesting that stimulus-focused processing may be greater for novel stimuli: For example, in a recent fMRI study, Poppenk, Moscovitch, & McIntosh (submitted) found greater activity for novel compared to repeated and previously known proverbs in brain regions responsible for perception and language. Also, Diana & Reder (2006) found worse recognition of pictures with low-frequency words superimposed on them than of pictures with high-frequency words superimposed on them, which they argued reflected greater distraction by the (relatively novel) low-frequency words, arising from greater processing requirements of those words. Consistent with this idea, they found in other experiments that recognition of low-frequency words suffered more from divided attention than did recognition of high-frequency words.

Another possible explanation of the familiarity bonus for source memory is that retrieved information from long-term memory (LTM) *facilitates binding of item and contextual information* by providing a scaffold onto which new memories can “stick”. This idea encompasses a family of mechanisms that we will refer to collectively as the *scaffolding* account. The key shared feature of all of these candidate mechanisms is that relevant prior experience leads to enhanced binding of item and context features. For example, one variant of the scaffolding hypothesis posits that associations that underlie stimulus representations in LTM permit familiar stimuli to be represented as a single “chunk” of information (Gobet et al., 2001), whereas novel materials must be represented as the combination of their features. To the extent that there are limits on the number of chunks that can be held in consciousness (Miller, 1956), the availability of more complex chunks for familiar items makes it possible to simultaneously represent a greater number of stimulus features for familiar vs. novel items. Putting these points together, if learning of item-context associations is focused on the contents of consciousness (Moscovitch, 2008), and participants can consciously represent more details of events involving familiar items (because of chunking), this implies that a richer web of item-context associations will be formed for familiar than novel items, leading to better source memory for familiar items at test (see the *Discussion* for other, related views). A closely related idea is that prior knowledge relating to an item promotes deeper, more elaborative processing of that item at study (Craik and Lockhart, 1972). These elaborations can then be linked to contextual features, leading (again) to a richer web of item-context associations than would otherwise be present.[1]

Crucially, the two families of explanations (attention and scaffolding) make different kinds of predictions. The key prediction of the attention hypothesis relates to the *overall amount* of contextual processing that is triggered by familiar vs. novel items: According to the attention

[1] In addition to increasing the number of encoded features, elaboration can also increase the *distinctiveness* of memory traces, which (in turn) will promote good source memory by reducing interference between stored memory traces. The Poppenk & Moscovitch (2011) dataset that we re-analyze in this paper used highly distinctive proverb stimuli; in this situation, we would expect the level of between-trace interference to be relatively low for novel proverbs, and it seems unlikely that familiarization would lead to a substantial reduction in the (already low) level of interference. However, in situations where stimuli are less distinctive – and, consequently, between-trace interference is more of a concern – this mechanism (i.e., prior exposure leading to increased distinctiveness and reduced interference) may have a strong effect on source memory performance.

hypothesis, there will be more contextual processing associated with familiar than novel items, since attention is directed towards stimulus processing to a greater extent for novel than familiar items. The scaffolding hypothesis, by contrast, does not make any special predictions about relative amounts of contextual processing that are triggered by familiar vs. novel items. Rather, the key prediction of the scaffolding hypothesis concerns the *relationship* between contextual processing at study and subsequent source memory. As noted above, processing of contextual features is not sufficient to yield good source memory; these features also need to be bound to the item representation. If – because of scaffolding – item-context binding is especially effective for familiar items, this implies that any factor that increases contextual processing will lead to a concomitant increase in source memory for these items. Conversely, if item-context binding is relatively *ineffective* for novel items, then it is possible that some factor might boost contextual processing but (because of poor binding) this increase might not lead to better source memory. Putting these points together: If the scaffolding hypothesis is true (such that retrieved LTM knowledge for familiar items facilitates binding), we would expect there to be a *tighter relationship* between contextual processing and source memory for familiar vs. novel items. As terminological shorthand, we will use the word "stickiness" to refer to the property of how easily contextual features are bound to item features; items that are more sticky will (by definition) show a stronger relationship between contextual processing at study and memory for item-context associations at test. Using this terminology, the key prediction of the scaffolding theory is that familiar items will be "stickier" than novel items.

To evaluate these hypotheses, we needed some way of measuring contextual processing (at study) that we could then relate to source memory accuracy (at test). To accomplish this goal, we took the data from a recently published fMRI study of source memory, and we used multi-voxel pattern analysis (MVPA; Norman, Polyn, Detre & Haxby, 2006) to obtain a covert, neural measure of the level of contextual processing at study. The specific dataset we investigated (Poppenk & Moscovitch, 2011) employed the same cognitive design originally used to disentangle familiarity and novelty effects (Poppenk, Köhler & Moscovitch, 2010a): Participants studied novel Asian proverbs, Asian proverbs that were repeated three times in a pre-study fMRI run, and previously known English proverbs while being scanned with fMRI. At study, participants performed one of two experimental tasks on each proverb. After all scanning was complete, participants were given a source memory test in which they indicated the task associated with each proverb at study. Using study-phase fMRI data, we trained a pattern classifier to measure the extent to which participants were engaged in each of the two experimental tasks at any given point in time: this assessment of on-task processing was the key measure of contextual processing that we used to test our hypotheses. To address the attention hypothesis, we compared the amount of on-task processing associated with novel, repeated and previously known proverbs. To address the scaffolding hypothesis, we compared the strength of the relationship between on-task processing and subsequent memory among novel, repeated and previously known proverbs, performing this comparison both within and between participants.

2. Methods

2.1 Dataset summary

The behavioural, data acquisition and preprocessing procedures associated with the dataset we reanalyzed are reported elsewhere (Poppenk & Moscovitch, 2011) but are summarized here for convenience. Briefly, eighteen right-handed young adults, all fluent in English, participated in the experiment (11 female; aged 21 to 34 years, mean age 26.1). The experiment consisted of three main phases (Table 1). Stimuli were 160 Asian proverbs (e.g., "A single hair can hide a mountain") and 80 English proverbs (e.g., "A stitch in time saves nine"). Please refer to Poppenk et al. (2010a) for a comprehensive list of these materials.

The mean number of characters in each proverb was 37.90 (SD=7.90; min=21; max=66). The mean number of words in each proverb was 7.26 (SD=1.72; min=4; max=14).

In Phase 1, participants were presented with 80 Asian proverbs three times while completing incidental study tasks involving these materials. The first time each proverb was presented, participants were required to interpret the meaning of each proverb (exposure time = 7.5 s; mean inter-stimulus interval, i.e., delay between stimulus offset and next stimulus onset = 5.8 s), responding with a button press when they arrived at a possible meaning. This task was scanned with fMRI. The second and third exposures of each proverb in Phase 1 were performed during acquisition of anatomical images; during these presentations, participants were required to guess whether proverbs were of Asian or South American origin (all were Asian).

During Phase 2, participants viewed the 80 proverbs that had been repeated three times in Phase 1 (*repeated* proverbs), another set of 80 Asian proverbs (not shown in Phase 1; *novel* proverbs), and 80 common English proverbs (*previously known* proverbs). The exposure time for each proverb was 4.5 s and the mean inter-stimulus interval was 4.8 s. For half (40) of the proverbs in each of these conditions, participants performed a *target-age* task: using a button-press, participants rated whether each proverb would be more suitable for an adolescent or an adult. For the other half of the proverbs in each list, participants performed a *quality-rating* task: using a button-press, participants decided whether each proverb was of good or poor quality. In Phase 2, proverbs were presented in blocks containing five proverbs of the same type (novel, repeated or previously known). To minimize the role of task switching in the experiment, these blocks were clustered into super-blocks where the encoding task was held constant. Each super-block was composed of three blocks (one novel, one repeated, and one previously-known block), the ordering of which was random. Encoding of the proverbs in Phases 1 and Phase 2 was incidental: participants were not warned about an upcoming memory test and, when asked (in a post-experiment debriefing session), none of the participants reported suspicion of a subsequent memory test.

Phase 3 took place at a computer following a 30-minute interval. Participants completed a source memory test that was based on the tasks performed in Phase 2: participants indicated whether they had performed the target-age task or quality-rating task on each proverb (with no time limit). After the source memory test, participants were shown all of the proverbs from the experiment and they were asked whether the proverb was known to them prior to the experiment – our assumption was that English proverbs would be pre-experimentally known and that Asian proverbs would not be pre-experimentally known. This was mostly but not completely true (known English proverbs: *M*=89.9%, *SD*=9.5%; known Asian proverbs: *M*=9.7%, *SD*=9.7%). Proverbs with prior knowledge responses inconsistent with the expected pattern were dropped from analysis. One participant was excluded from the experiment for head motion during scanning and another for getting all items correct in the Phase 3 source memory task (to conduct our analysis, we required some trials where the source was correctly remembered and some trials where the source was incorrectly remembered). In total, 16 participants were included.

Scanning was performed using a 3 Tesla whole-body MRI system (Siemens, Erlangen, Germany) installed at Baycrest Hospital in Toronto, Canada. T1-weighted high-resolution MRI volumes were collected using a standard 3D MPRAGE pulse sequence, with slices collected in an oblique axial orientation (160 axial slices; FOV = 256 mm; 256 × 192 matrix; TR = 2000 ms; TE = 2.63 ms; flip angle = 9°). The Phase 1 task that was scanned using fMRI was completed in two 375 volume-runs collected using T2-weighted echo-planar image (EPI) acquisition (24 axial oblique slices; FOV = 200 mm; 64 × 64 matrix; TR = 1500 ms; TE = 30 ms; flip angle = 60°; no parallel acquisition). Phase 2 was completed in four 485-volume runs collected using the same sequence. All procedures were approved by the research ethics boards at Baycrest and the University of Toronto.

Initial preprocessing of the T2-weighted functional images was performed using FSL (FMRIB Software Library; Smith et al., 2004). Following slice timing and motion correction within each image series, high-pass filtering was applied (sigma = 49.5 s). The series was co-registered to the participant's high-resolution T1-weighted anatomical image, which was used as a reference to transform functional data into standardized MNI space at a voxel resolution of 4 × 4 × 4 mm^3. Normalized functional images were entered into a semi-automated probabilistic independent component analysis to filter residual motion artifacts, high-frequency scanner noise and artifacts related to gradient timing errors (Beckmann & Smith, 2004). High-amplitude spike volumes were dropped. Mean white matter intensity for each volume was calculated using a mask constructed using a probabilistic atlas (Mazziotta et al., 2001) and was residualized from the image series as a global intensity normalization step.

2.2 Additional preprocessing

Cortical segmentation of the high-resolution anatomical images was performed in a semi-automated fashion using the Freesurfer image analysis suite, which is documented and available online (http://surfer.nmr.mgh.harvard.edu). The technical details of these procedures are described elsewhere (e.g., Fischl et al., 2004). Briefly, this processing includes removal of non-brain tissue using a hybrid watershed/surface deformation procedure, automated Talairach transformation, intensity normalization, tessellation of the gray matter white matter boundary, automated topology correction and surface deformation following intensity gradients, parcellation of cortex into units based on gyral and sulcal structure, and creation of a variety of surface based data including maps of curvature and sulcal depth. Freesurfer morphometric procedures have been demonstrated to show good test-retest reliability across scanner manufacturers and across field strengths (Han et al., 2006). Manual quality control checks were performed during the procedure. We assembled all cortical and subcortical grey matter segmentations into a mask for identifying voxels to be used in our analysis, omitting motor, premotor and supplementary motor regions to discourage classification from depending on distinctive motor response effects in individual conditions. This mask was resampled into the space of our functional images and used to inclusively mask the functional data, allowing only data from grey-matter voxels outside of motor regions into our analysis. All volumes were concatenated, and signal from each voxel was z-scored within run (this was done separately for Phase 1 and Phase 2 data).

2.3 Classifier training

To address our hypotheses, it was necessary to obtain a covert, ongoing measure of participants' engagement in the Phase 2 encoding tasks. We derived such a measure using the fMRI data collected during Phase 2 by applying the following two-step process: First, we trained a classifier to be sensitive to the features of the neuroimaging data that distinguished between the two encoding tasks. Second, we used the classifier to measure (over time) how engaged participants were in the appropriate encoding task (i.e., the encoding task that they were instructed to perform on that item).

We conducted our classifier analysis in Matlab using functions from the Princeton MVPA Toolbox (Detre et al., 2006; available for download at http://www.pni.princeton.edu/mvpa/; see also Norman et al., 2006, for a discussion of the logic and affordances of MVPA}. Classifier training was performed separately for each participant using fMRI data collected during the two incidental encoding tasks, which were collected using a mixed design. Our specific goal was to train the classifier to track the extent to which participants were engaged in the target-age task vs. the quality-rating task. To support this training, we created two regressors. The first identified all time points where instructions to complete the target-age task were active, and the second identified all time points where instructions to complete the quality-rating task were active. As events were organized into blocks of like events (same task and novelty condition), we averaged the volumes within each block into a single volume, first shifting regressors by three volumes (i.e., 4.5 s) to approximate hemodynamic lag effects associated with the blood-oxygen level dependent response in fMRI data.

To avoid circularity when estimating cognitive state information with a classifier, it is essential to use different portions of the data for classifier training and for classifier testing (Kriegeskorte, Simmons, Bellgowan & Baker, 2009). Towards this end, we used a cross-validation scheme in which three of the four Phase 2 runs were allocated to model training (i.e., the training set) and the omitted run was allocated to testing (i.e., the testing set). As there were four possible runs to omit, this scheme involved creating four cross-validation folds, each with a different combination of training and testing sets. For technical reasons, several participants had only three runs of data available; accordingly, three cross-validation folds were prepared for those individuals.

To pre-select voxels likely to be most useful for our classifier, we performed a non-parametric bootstrap contrast of the mean functional image associated with each task condition as a feature selection step (this feature selection procedure was performed separately on the training data within each of the cross-validation folds). In this analysis, for each task condition, we computed the voxel-wise mean signal within the training set. We then computed the difference between conditions. We randomly sampled with replacement from the n blocks available for training until n samples had been obtained, and calculated condition means in this manner 100 times. To estimate standard error of the mean, we computed the standard deviation of the difference between conditions for each of the 100 bootstrap samples. Finally, we expressed the original contrast of condition means as a ratio over standard error (i.e., bootstrap ratio or BSR), a metric approximately equivalent to a z score where the bootstrap distribution is normal (Efron & Tibshirani, 1986). The final product of feature-selection was a voxel map, computed for each cross-validation fold, describing the reliability of the difference between the two task conditions in each voxel. We sorted each map by these values to identify, in descending order, the voxels most likely to be informative for separating the conditions.

Classifier training was conducted using a ridge regression algorithm (as implemented in the Princeton MVPA Toolbox). Ridge regression learns a β weight for each input feature (voxel) and uses the weighted sum of voxel activation values to predict outcomes (in this case, a binary vector indicating which task is associated with each volume). The ridge regression algorithm optimizes each β to simultaneously minimize both the sum of the squared prediction error across the training set and also the sum of the squared β weights (technical details are described elsewhere; see Hastie, Tibshirani, & Friedman, 2001, and Hoerl & Kennard, 1970). A regularization parameter (λ) determines how strongly the classifier is biased towards solutions with a low sum of squared β weights; when this parameter is set to zero, ridge regression becomes identical to multiple linear regression. The solution found by the classifier corresponded to a β map for each regressor describing the spatial pattern that best distinguished that regressor's condition from other conditions (with regularization applied). We selected a ridge regression algorithm (as opposed to a classifier with a nonlinear transfer function like logistic regression) because of its sensitivity to intermediate activation values – we wanted to be able to track small fluctuations in on-task processing on a trial-by-trial basis.

To set values for the ridge regression penalty parameter (λ) and the number of voxels that were provided to the classifier for training (*m*), we conducted a grid search over the space of these two parameters; for λ we examined values in the set (0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, and 100) and for *m* we examined values from 500 to 9000 in increments of 500. For a given value of *m,* we selected the most informative *m* voxels, as identified using the feature-selection procedure described above. For each set of parameters, we computed the grand mean of classification accuracy across cross-validation folds and individuals; for this analysis, classification accuracy was defined as the proportion of blocks where the classifier output for the correct task was greater than the classifier output for the incorrect task. Figure S1 shows the results of the grid search; the parameters yielding maximum classification accuracy were $m = 7000$ voxels and $\lambda = 10$ (with these parameters, accuracy = 0.58; chance = 0.50). These parameters were used for all of the subsequent classification analyses. Crucially, the hypotheses we investigated in our analysis did not concern the overall level of classification accuracy; rather, our hypotheses concerned differences in classifier output across experimental conditions. Because we optimized λ and *m* for overall accuracy (across conditions) rather than differences in classifier output across conditions, this optimization procedure did not introduce circularity into our hypothesis testing.

To gain insight into which brain regions were driving classifier performance, we constructed *importance maps* for the target-age classifier and the quality-rating classifier using the procedure described in (McDuff, Frankel, & Norman, 2009). This procedure identifies which voxels were most important in driving the classifier's output when that task was present. For each task, we computed the average activation of each voxel when that task was present (note that voxel time courses were z-scored within runs, according to the procedure described above). Voxels with a positive β weight and a positive z-scored average activation value were assigned a *positive* importance weight for that task, with a value equal to the product of the weight and the activation value. Voxels with a negative β weight and a negative z-scored average activation value were assigned a *negative* importance weight for that task, with a value equal to the product of the weight and the activation value. Voxels where the sign of the β weight differed from the sign of the average activation value were assigned an importance value of zero. Note that, with these equations, both positive and negative importance values indicate a net positive contribution of that voxel to activating the task classifier (when that task is present). The absolute value of the importance score indicates the size of that voxel's contribution. The sign of the importance value indicates whether the voxel contributes via a characteristic deactivation that is picked up by the classifier (via a negative weight), or a characteristic activation that is picked up by the classifier (via a positive weight). We computed these importance values for each cross-validation fold within a participant and then averaged the values together. We then averaged together the participant-specific importance maps to get a group importance map. For further details on the logic on this procedure, please refer to McDuff et al. (2009). Note that importance maps do not provide a comprehensive treatment of where task-relevant information is located in the brain; as discussed by Norman et al. (2006), there are many reasons why informative voxels might be assigned zero weights. The importance maps are presented in Figure S2 for informational purposes only, and no inferential statistics were computed based on these maps.

2.4 Classifier output as a dependent measure

Having successfully trained a classifier to discriminate between neural patterns associated with the two experimental tasks, our next step was to use this classifier to detect fluctuations in on-task processing over time (i.e., fluctuations in the degree to which participants were performing the instructed encoding task). To obtain a temporal "read-out" from our ridge regression classifier corresponding to fluctuations in on-task processing, functional volumes from individual time points were evaluated. This evaluation was conducted in concordance with the cross-validation regime described above (i.e., to avoid circularity, each run was evaluated using a classifier that was trained on the other runs; classifier outputs from the different runs were then reassembled into a single time series). The result of this processing was two time series: a time series indicating TR-by-TR (i.e., one 1.5 s functional scan at a time) fluctuations in target-age task activity, and a time series indicating TR-by-TR fluctuations in quality-rating task activity. To facilitate interpretation of this output, we computed (for each TR) the difference between classifier output related to the correct task and classifier output related to the incorrect task (e.g., if a particular time point was labeled as a "target-age" time point, we computed the output of the target-age classifier minus the output of the quality-rating classifier).

At this stage of our analysis, the process of converting our Phase 2 fMRI time series into an *on-task processing* time series was complete. By providing an estimate of the extent to which participants were engaged in the experimental tasks at any given point in time during incidental encoding in Phase 2 (as opposed to focusing on stimulus processing, the wrong task, daydreaming, or other non-task operations), this new time series fulfilled our need for a measure of trial-by-trial fluctuations in on-task processing during Phase 2. Our next task was to analyze how on-task processing (measured during Phase 2) varied as a function of 1) stimulus novelty and 2) whether the item's source (i.e., the task that was performed during Phase 2) was subsequently remembered correctly during Phase 3. Whereas the attention hypothesis makes no particular predictions about the relationship between Phase 2 on-task processing and Phase 3 source memory accuracy (other than that it should be positive), the scaffolding account predicts that the relationship should be stronger for familiar than novel items.

As a first step towards understanding the relationship among on-task processing, novelty, and memory, we parsed the on-task processing time series in an event-locked fashion, extracting the series of eight values that began with each event onset (we refer to these time points as TR 0 through TR 7). This corresponded to a window size of 12.0 s, which captured the 4.5 s of stimulus exposure and 4.8 s average inter-stimulus interval that followed while allowing for a delayed hemodynamic response. Note that, due to the study's use of an event-related design with a relatively short ISI, the hemodynamic responses to successive events overlapped in time to some degree. If the degree of on-task processing differs from trial to trial, carryover of the hemodynamic response for the previous trial will serve as a source of noise when classifying the current trial; these carryover effects will make it harder to observe a relationship between contextual processing on a particular trial and subsequent memory. As discussed in the *Results* section below, we managed to observe a significant relationship between Phase 2 contextual processing (as measured by the classifier) and Phase 3 source memory despite this handicap.

Finally, we organized event responses according to novelty condition (novel, repeated and previously known). We also organized responses according to whether the participant subsequently responded correctly to the item on the Phase 3 source memory test (*correct source* or *incorrect source*). Once events were sorted in this fashion, we either took the average of all events split by novelty condition (novel, repeated and previously known) and time point (eight TRs post-stimulus-onset) or we took the average of all events split by novelty condition, subsequent memory in Phase 3, and time point.

2.5 Trial-sorting using Phase 1 data

By comparing the on-task processing time series from Phase 2 that had been sorted according to novelty and memory success, we were able to investigate questions concerning how contextual processing and "stickiness" (operationalized as the strength of the relationship between contextual processing in Phase 2 and subsequent source memory in Phase 3) may vary as a function of stimulus novelty. As a further, more detailed test of the scaffolding model's prediction that prior episodic encoding should boost the stickiness of an item, we next sought to investigate whether items within the repeated condition themselves varied in stickiness according to how well they were encoded when they were presented during the Phase 1 pre-study period. Based on prior data showing that hippocampal activity predicts subsequent source memory (Davachi, Mitchell, & Wagner, 2003; Stark & Okado, 2003), we used hippocampal activity during Phase 1 as a neural measure of episodic encoding strength. Specifically, we used an anatomical segmentation of the hippocampus derived from our high-resolution MRI scans (described above) to inclusively mask and extract the mean signal of the hippocampus in each hemisphere. Because these signals were highly correlated (mean $r = 0.77$), we merged them into a single time series of Phase 1 mean hippocampal activity to facilitate interpretation. To capture the hippocampal response to each proverb, we parsed the Phase 1 fMRI time series in an event-locked fashion in the manner described above (i.e., using a window size of 12.0 s, starting with the onset of the proverb). We averaged across all items to locate the peak hippocampal response, which at the group level took place 9.0–10.5 s after stimulus onset; for each item, we recorded the value of the hippocampal time series at this time point and used this as our neural measure of episodic encoding.

To assess how episodic encoding strength during Phase 1 affects stickiness, we ran a median split analysis: Within each participant, we divided proverbs into two sets based on whether their Phase 1 hippocampal activity value was above or below the median hippocampal activity value for that participant. We then separately analyzed the “above-median” and “below-median” items. First, we separately computed levels of on-task processing during Phase 2 for above-median and below-median items. Next, we subdivided these groups by subsequent memory during Phase 3 (source correct or incorrect), yielding four on-task processing time series (above-median source correct and incorrect; and below median source correct and incorrect). If episodic encoding strength during Phase 1 boosts stickiness, we would expect to see a stronger relationship between on-task processing and subsequent memory for “above-median” items (where episodic encoding strength was relatively high) than for “below-median” items (where episodic encoding strength was relatively low).

2.6 Within-subjects comparisons

To provide a statistical test of possible on-task processing differences arising within participants (i.e., based on novelty conditions, Phase 3 source memory success and Phase 1 hippocampal signal), group-level pairwise analyses between condition means were conducted using a non-parametric bootstrapping analysis. At each time point, pairwise differences between condition means across participants were calculated. These computations were repeated 100 times, each time drawing n samples with replacement from the group of n participants. The standard deviation of differences provided a standard error estimate for each comparison. We divided the overall mean difference by the difference standard error derived from bootstrap resampling to obtain a BSR, which can be treated as an approximate z statistic (Efron & Tibshirani, 1986). We set our significance threshold at an absolute value of BSR 1.96 (approximately corresponding to a 95% confidence interval).

2.7 Between-subjects correlations

In addition to exploring how average levels of on-task processing varied as a function of novelty condition and subsequent memory, we explored the relationship across participants between levels of on-task processing and source memory performance. This is an alternative way of testing the scaffolding hypothesis: The “stickier” that items are, the greater the correlation should be (across participants) between the level of on-task processing and the level of subsequent memory performance.

For each of the three novelty conditions (novel, repeated, previously known), we computed, for each participant, the average level of Phase 3 source memory accuracy and also the average level of Phase 2 on-task processing at TR4 (this was the time point at which, in all conditions, we observed reliable on-task processing; see the *Results* section for details). Next, for each condition, we computed the Pearson correlation (across participants) between on-task processing with source memory accuracy.

To calculate the reliability of each correlation, we collected 1000 bootstrap samples, each time selecting with replacement a different subset of *n* individuals from our pool of *n* participants. For each sample, we computed the correlation between the predictor and predicted variables. The upper limit *(ul*) for an *x*% confidence interval was set to the value of the Pearson correlation in percentile *x* of the bootstrap distribution; the lower limit *(ll*) for the confidence interval was set to the value of the beta score in percentile 100-*x* of this distribution. Confidence intervals that did not encompass zero were considered reliable at the given level of confidence.

We were also interested in computing whether the values of these correlations (computed within condition) differed across conditions; in particular, we wanted to know whether the correlation was higher in the familiar-proverb conditions (repeated and previously known) than the novel-proverb condition. If found, this pattern would be indicative of increased stickiness for familiar vs. novel proverbs. We computed r_{diff}, upper confidence limits (ul_{diff}) and lower confidence limits (ll_{diff}) for the difference between two correlations using the following formulae proposed by Zou (2007) for application to two bootstrapped correlation confidence intervals:

$$\begin{aligned} r_{diff} &= r_1 - r_2 \\ ll_{diff} &= r_1 - r_2 - \sqrt{(r_1 - ll_1)^2 + (ul_2 - r_2)^2} \\ ul_{diff} &= r_1 - r_2 + \sqrt{(ul_1 - r_1)^2 + (r_2 - ll_2)^2} \end{aligned}$$

3. Results

3.1 Behavioural results

Figure 1 shows the behavioural source memory results from Phase 3. There was a main effect of novelty, BSR = 3.23, $P < 0.005$. Post-hoc tests revealed better source memory for familiar proverbs than novel ones – this held true both for repeated proverbs (repeated > novel, BSR = 3.47, $P < 0.001$) and previously known proverbs (previously known > novel, BSR = 6.06, $P < 0.001$). The mean reaction times were 3110 ms (SD = 385 ms) for novel proverbs; 2547 ms (SD = 420 ms) for repeated proverbs; and 2372 ms (SD = 423 ms) for previously known proverbs.

3.2 fMRI classification results

3.2.1 On-task processing as a function of time—Our Phase 2 on-task processing measure consisted of the difference between the classifier output for the correct task and the classifier output for the incorrect task. This measure provided us with a quantitative estimate of the amount of task-specific processing that was present at each time point. Figure 2 shows event-locked averages of on-task processing for the novel, repeated, and previously-known proverbs. Our first analysis of on-task processing investigated when (in the event-locked time course) levels associated with novel, repeated and previously known proverbs could be distinguished from chance (where chance = 0). As shown in Table 2, on-task processing was reliably positive in all conditions for at least one time point following stimulus offset; in particular, above-zero signal in Phase 2 on-task processing was present in one or more conditions at post-onset TRs 3, 4 and 5 (i.e., the time window of 4.5 to 9.0 s post-onset), approximately aligning with the hemodynamic peak that could be expected based on stimulus presentation from 0.0 s to 4.5 s.[2] To limit the total number of comparisons performed, we restricted subsequent group statistical testing to these time points, with particular emphasis on TR 4, the one time point at which on-task processing was observed in all conditions.

3.2.2 On-task processing as a function of condition—Next, we evaluated whether levels of on-task processing were different for novel vs. familiar items. As shown in Table 3 (under “Phase 2 on-task processing”), we observed significantly greater levels of on-task processing for previously known proverbs than novel proverbs at TR 4. On-task processing for previously known proverbs was also significantly greater than on-task processing for repeated proverbs at TRs 3 and 4. Levels of on-task processing did not significantly differ between repeated and novel proverbs at any of the time points.

3.2.3 Relationship between on-task processing and subsequent memory as a function of condition—Figure 3 (parts A–C) shows levels of on-task processing split by condition (novel, repeated, previously known) and subsequent source memory (correct, incorrect). Figure 3, part D re-plots the results from parts A–C as subsequent memory effects (i.e., the difference in on-task processing for items based on whether source was subsequently remembered correctly or incorrectly), as a function of condition. As shown in Table 3 (under “Phase 2 on-task processing subsequent memory effect”), the subsequent memory effect was significantly larger for repeated than novel items at TR 4. None of the other differences were statistically reliable (note that there was a trend for a larger subsequent memory effect for repeated than for previously known proverbs at TR 4, but this trend did not meet our criteria for significance).

3.2.4 Median split based on Phase 1 hippocampal activity—To further assess how prior episodic encoding in the repeated-proverb condition affected processing during Phase 2, we further sorted the repeated items according to whether they were associated with above-median or below-median hippocampal signal during Phase 1. Figure 4 (part A) shows the average level of Phase 2 on-task processing for above-median and below-median items. As shown in Table 4 (under “Phase 2 on-task processing”), overall levels of on-task processing did not differ significantly for above-median vs. below-median items at any of the time points. Figure 4 (parts B and C) shows on-task processing split by above/below median and subsequent source memory. Figure 4, part D re-plots the results from parts B and C as subsequent memory effects. As shown in Table 4 (under “Phase 2 on-task processing subsequent memory effect”), the subsequent memory effect was significantly larger for above-median items than for below-median items at TR 4.

3.2.5 Between-subjects correlations between on-task processing in Phase 2 and source memory in Phase 3—As a further test of the memory scaffolding hypothesis, we computed – for each condition -- the correlation (across participants) between the average level of on-task processing at TR4 (the time point at which reliable on-

[2] The finding of above-chance task classification for novel items during TR0 and TR1 is attributable to the fact that trials were blocked by task during this phase. As the average inter-trial interval was 9.2s (including approximately three TRs of stimulus presentation and three TRs of fixation), TR0 and TR1 correspond to, on average, TR6 and TR7 of the previous trial. Therefore, good classification at this early time point is highly likely to reflect carry-over from the previous trial, which, in 93% of events, used the same encoding task as the current trial.

task processing was observed in all conditions; see Section 3.2.1) and source memory accuracy. We were interested in whether the strength of this relationship (between Phase 2 on-task processing and Phase 3 source memory accuracy) might differ as a function of novelty condition. Figure 5 shows the results of this correlation analysis. Part A of the figure shows the correlation between on-task processing and source memory accuracy when we collapse across novelty conditions. Overall memory performance was reliably predicted by overall Phase 2 on-task processing, $r = 0.28$, one-way $P < 0.05$, 90% $ll = 0.03$, $ul = 0.51$. Figure 5, parts B–D show the correlation between on-task processing and source memory accuracy for each of the novelty conditions. This predictive relationship was stronger for repeated proverbs than novel proverbs, $r_{\text{diff}} = 0.33$, one-way $P < 0.05$, 90% $ll_{\text{diff}} = 0.01$, $ul_{\text{diff}} = 0.69$, and was also stronger for repeated proverbs than previously known proverbs, $r_{\text{diff}} = 0.46$, one-way $P < 0.05$, 90% $ll_{\text{diff}} = 0.05$, $ul_{\text{diff}} = 0.75$. No difference was observed between previously known and novel proverbs, $r_{\text{diff}} = -0.13$, 90% $ll_{\text{diff}} = -0.43$, $ul_{\text{diff}} = 0.34$.

4. Discussion

In our study, we sought to evaluate two candidate hypotheses (the *attention hypothesis* and the *scaffolding hypothesis*) that might be able to explain recent observations (also confirmed here) of superior source memory for repeated and previously known items over novel items (Poppenk et al., 2010a). According to the attention hypothesis, retrieval of stored memories in response to familiar proverbs should mitigate the need to spend cognitive resources on stimulus-related processing, thereby allowing participants to engage more deeply in task-related processing. According to the scaffolding hypothesis, even when overall levels of on-task processing are similar, retrieved memories can facilitate memory encoding for familiar items by providing a scaffold onto which new memories can “stick”. This scaffold should boost the efficiency with which item-context associations are formed, resulting in a stronger relationship between the degree of on-task processing observed and subsequent memory.

To evaluate these hypotheses, we employed a covert neural measure of on-task processing, obtained using a pattern classifier applied to our Phase 2 fMRI memory encoding data, which provided us with a window into participants’ cognitive state throughout the encoding phase of the experiment. The classifier measured participants’ engagement in study tasks during Phase 2 of our study. In keeping with the encoding specificity principle (whereby the type of information processing at encoding determines the type of information available in memory at test; Tulving & Thomson, 1973), we found that elevated signal in our temporal read-out of on-task processing was associated with greater subsequent memory for task-related source information. That this basic prediction was met suggested that our on-task processing classifier measured its intended construct with a sufficient level of signal to resolve cognitive effects of interest.

We applied our on-task processing measure to test the predictions of the attention hypothesis and the scaffolding hypothesis. To test the attention hypothesis, we evaluated whether overall levels of Phase 2 on-task processing were higher for familiar than novel items. We observed this pattern for previously known proverbs, where there was greater on-task neural processing relative to novel proverbs, but no such increase was present for repeated proverbs. To test the scaffolding hypothesis, we looked at whether the relationship between Phase 2 on-task processing and Phase 3 source memory was stronger for familiar vs. novel items. We observed this pattern in the repeated condition, where there was significantly greater stickiness relative to novel proverbs; numerically, there was greater stickiness for previously known proverbs than for novel proverbs, but this difference was not significant. Evidence from our between-subjects correlation analyses yielded a similar pattern of results: Across participants, the correlation between on-task processing and source memory was reliably larger for repeated proverbs than for previously known and novel proverbs, and the correlations associated with previously known and novel proverbs did not differ from one another.

That repeated items were found to be more "sticky" than previously known ones in our between-subjects correlation analysis (with a non-significant trend in our within-subjects subsequent memory analysis) raises the question of why different mechanisms should apply when both repeated and previously known proverbs have a pre-existing representation in LTM. There are two obvious differences between the two types of proverbs. First, through years of experience and consolidation time, previously known items are much more strongly represented in LTM than repeated proverbs. Second, at the start of Phase 2, repeated proverbs are very likely to be associated with a stored episodic memory trace by virtue of having been recently presented in Phase 1, whereas the episodic memory trace associated with previously known proverbs may have long since decayed from episodic memory (Moscovitch et al., 2005; Squire & Zola, 1998), given the low frequency of proverb use in English discourse. Instead, previously known proverbs may persist only in semantic memory, a memory store that contains general facts and knowledge rather than detailed contextual features of specific events (Tulving, 1972). The finding of greater stickiness overall for repeated proverbs (which have an episodic trace from Phase 1) compared to novel and previously known proverbs (which do not have an episodic trace from phase 1) suggests that an important factor in a proverb's stickiness may be the advance presence of an intact episodic memory trace that contains the proverb. To test this idea, we conducted a median split analysis on data from the repeated-proverb condition. Drawing upon evidence that hippocampal activity at encoding is predictive of subsequent episodic memory retrieval (Davachi et al., 2003; Stark & Okado, 2003), we used mean hippocampal activity elicited by the proverb in Phase 1 (i.e., the pre-study repetition phase) as a neural index of episodic encoding. The results of this analysis supported our hypothesis: In keeping with the idea that successful episodic encoding during Phase 1 makes memories more sticky, on-task processing was a better memory predictor for proverbs with above-median hippocampal activation during Phase 1 than for proverbs with below-median hippocampal activation.

The idea that retrieval of previously-formed episodic memories potentiates subsequent association formation is further substantiated by previous cognitive and neural findings. For example, Wichawut and Martin (1971) found that the probability of cued recall of word associates (A–C) covaried positively with the degree to which other associates were linked to the same cues (A–B) during earlier training. However, while cognitive studies of this sort are consistent with the idea that episodic memories can provide scaffolding, they do not speak to exactly how episodic memories are supporting subsequent learning, nor do they indicate exactly which neural mechanisms are most important for these effects. These factors have been better addressed by investigation of fear conditioning in rodents: Numerous studies have found that familiarity with an environment greatly facilitates learning of an association between the environment and electric shocks. This is referred to in the literature as the *immediate shock effect* (Fanselow, 1990; Rudy, Barrientos, & O'Reilly, 2002). Importantly, Rudy and colleagues found that the facilitatory effect of prior experience on learning depends critically on the hippocampus: even though contextual fear learning without prior environmental exposure was preserved among hippocampally-lesioned rats in their study, fear memory was no longer enhanced by prior environmental exposure as it was in sham-lesioned rats (i.e., the immediate shock effect was lost).

Rudy and O'Reilly (2001) explain the immediate shock effect in terms of hippocampally-mediated pattern completion: when an animal initially explores the environment, it forms a hippocampally-mediated *conjunctive encoding* of all of the features of the environment; after this conjunctive memory is formed, thinking of one feature of the environment will automatically trigger hippocampal pattern completion, thereby retrieving other features of the environment. Rudy and O'Reilly's account of the immediate shock effect rests on two key claims. The first claim is that, when shock occurs, conditioning occurs to all features of the environment that were *actively represented* in the animal's brain at the moment of the shock, consistent with the analogous proposal in humans that episodic memory encoding specifically captures the contents of consciousness (Moscovitch, 2008). The second claim is that, because hippocampal pattern completion allows animals to represent both perceived and remembered features rather than perceived features alone, animals with an intact hippocampus and prior experience with the environment will actively represent a larger number of features than animals with a lesioned hippocampus. Putting these claims together: if animals with an intact hippocampus and prior experience actively represent a larger number features at the moment of shock, conditioning will occur to a larger number of features, resulting in a stronger overall level of conditioned fear. Rudy and O'Reilly (1999) present additional converging evidence for this "conjunctive encoding" account: They found that, when the testing environment is presented to the animal in a piecewise fashion during the familiarization phase (thereby preventing the animal from forming a conjunctive encoding that binds together the different pieces of the environment), the beneficial effect of familiarization on learning is greatly diminished.

While there are numerous differences between fear conditioning and our paradigm, it is easy to see how the basic logic would apply to our results: When participants first encounter a novel Asian proverb in Phase 1, they will think about many different aspects of the proverb's meaning. The hippocampus will form a conjunctive encoding that binds these different meaning features together. Later, when the proverb appears during Phase 2, thinking about one meaning feature will trigger pattern completion of other meaning features. As a result of this pattern completion, participants will actively represent a larger number of stimulus features for repeated proverbs compared to proverbs that are being presented for the first time in Phase 2. To the extent that associative learning only occurs for actively represented features (as proposed by Moscovitch, 2008) and that more features are actively represented for familiar materials (as proposed by Miller, 1956), more item-context learning should occur during Phase 2 for familiar than novel proverbs.

An important limitation of this theory is that it does not explain the absence of enhanced stickiness for previously known over novel proverbs. While previously known proverbs may not trigger retrieval of episodic memories, they have stored representations in semantic memory that should evoke a similar kind of pattern completion process that (in turn) should have increased the size of the observed perceptual set. Along these lines, some of the most powerful early demonstrations of chunking involved links between increased perceptual set size and expertise (Chase & Simon, 1973; Ericsson & Kintsch, 1995; Miller, 1956), suggesting that conjunctive encoding effects based on prior knowledge should be as prominent as effects based on repetition, if not more so. Furthermore, Tulving and Markowitsch (1998) specifically argued that new episodic memories require a foundation in semantic memory before they can become established.

One possible explanation for our unexpected result is that levels of on-task processing for previously known proverbs were effectively at ceiling, thereby eliminating poor contextual encoding as a source of memory failures. If the residual source memory errors that did occur for previously known proverbs were primarily driven by retrieval factors such as misattribution or bias (Schacter, 1999) instead of factors at encoding, this would explain why our Phase 2 classifier metric was not especially diagnostic of subsequent memory success. An alternative possibility is that there is something special about hippocampally-based pattern completion effects: for instance, item details that are pattern-completed by the hippocampus may be more easily incorporated into subsequent episodic associations (e.g., with task information) than details that are pattern-completed by cortex, perhaps because the original memory trace remains labile (for discussion, see Nader, 2003). The dataset described here does not arbitrate between these (and other) possibilities.

As a limiting factor, it is worth noting that the dataset described here utilized only semantically rich verbal stimuli, leaving open the possibility that the effects described here are limited to this stimulus domain. However, a repetition advantage in source memory has also been observed using natural scenes (Poppenk et al., 2010b) and nonwords and unfamiliar faces (Lee et al., 2012). At this point, it is not yet clear whether these apparently similar effects were driven by increased attention or increased stickiness; the families of mechanisms that we evaluated in the current paper do not specifically predict that different mechanisms would apply on the basis of verbal versus nonverbal stimulus materials.

It is also worth noting that Lee et al. (2012) ran a version of their experiment using pre-experimentally familiar stimuli (common words and famous faces); in this condition, they found a repetition *disadvantage* for source memory (i.e., worse source memory for repeated compared to non-repeated stimuli). Likewise, Kim et al. (2012) found that repetition impaired source memory when simple drawings of common objects were used at study. Taken together with the findings (reviewed above) showing a repetition advantage for novel stimuli, the Kim et al. (2012) and Lee et al. (2012) findings suggest that repetition may involve costs as well as benefits. As discussed throughout this paper, repetition may be beneficial in providing participants with an advance opportunity to create LTM representations of novel stimuli (e.g., unfamiliar proverbs, non-words or faces) that subsequently provide “scaffolding” for new source memories. However, there are also numerous reasons why repetition could hurt source memory: for example, Kim et al. (2012) discuss how participants may “tune out” the perceptual features of repeated stimuli, leading to poor encoding of item-context associations. For stimuli without pre-existing LTM item representations, the benefits of repetition (in terms of increased scaffolding) may outweigh the costs. However, for stimuli with LTM item representations that are already well-established (e.g., previously known objects, words or faces), the costs of repetition may dominate, since the scaffolding for these items is already present and further repetitions may not substantially boost the item’s stickiness. In this respect, it would be diagnostic to explore how repetition affects source memory for previously known proverbs in our paradigm. According to the “costs and benefits” idea outlined above, repetition of previously known proverbs may actually harm source memory for these proverbs.

Having said this, there are reasons to think that the repetition disadvantage observed by Kim et al. (2012) and Lee et al. (2012) may not generalize to our paradigm. Importantly, the Kim et al. (2012) and Lee et al. (2012) studies that found worse source memory for repeated items tested memory for perceptual source features (memory for location and background color), whereas our source judgment probed memory for more “reflective” features (i.e., task: rating the quality of the proverb vs. rating the target age for the proverb). Kim et al. (2012) explain their findings by arguing that repetition can cause a general shift in processing from externally-focused perceptual processing to internally-focused reflective processing, thereby impairing memory for the perceptual source features (location, background color) used in their study (see Chun & Johnson, 2011, for further discussion of perceptual/reflective tradeoffs). To the extent that our source features are reflective, this perceptual-to-reflective shift (if it occurs) may actually help source memory in our paradigm.

In conclusion, we found evidence that both attention and scaffolding mechanisms may be useful for explaining advantages of prior knowledge and repeated materials over novel ones in source memory tests. Unexpectedly, we also found evidence hinting at a dissociation, whereby the attention mechanism appears to benefit previously known items (more so than repeated items), and the scaffolding mechanism appears to benefit repeated items (more so than previously-known items, although this difference was not always significant). Because the attention mechanism could only be observed reliably for previously known proverbs, we hypothesize that superior memory in those conditions was supported by retrieval of stimulus information from semantic memory that reduced stimulus processing demands and enabled greater attention to contextual processing. Because the scaffolding mechanism could only be observed reliably in conditions where there was likely to be a strong pre-existing hippocampal memory trace (i.e., excluding previously known proverbs, novel proverbs, and repeated proverbs that evoked a weak hippocampal signal during Phase 1), we hypothesize that superior memory in those conditions was supported by hippocampal conjunctive encoding during Phase 1 and pattern-completion (during Phase 2) of the conjunctive hippocampal trace. However, the possible relevance of scaffolding mechanisms to previously known proverbs should not be ruled out; as noted above, a ceiling effect on levels of on-task processing may have obscured the relationship between on-task processing and (subsequent) source memory in this condition.

Speaking generally, our results provide an illustration of how pattern classifiers can be used to evaluate cognitive theories of memory. It should be acknowledged that the experiment described here was not specifically designed to test the attention and scaffolding hypotheses, and our results involving possible dissociations between previously known proverbs and repeated proverbs were not predicted ahead of time. Nonetheless, the current study provides grounds for thinking that different mechanisms may support superior memory for previously known and repeated proverbs. Additional focused experimentation will be needed to substantiate these claims.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

J.P. was supported by an NSERC postdoctoral fellowship and NIH R01 MH069456 to K.A.N. We thank A. Conway for helpful statistical advice and M. Moscovitch and A.R. McIntosh for granting permission to reanalyze the dataset examined in this paper.

References

Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Transactions on Medical Imaging. 2004; 23:137–152. [PubMed: 14964560]

Castel AD, Craik FI. The effects of aging and divided attention on memory for item and associative information. Psychology and Aging. 2003; 18:873–885. [PubMed: 14692872]

Chase WG, Simon HA. Perception in chess. Cognitive Psychology. 1973; 4:55–81.

Chun MM, Johnson MK. Memory: enduring traces of perceptual and reflective attention. Neuron. 2011; 72:520–535. [PubMed: 22099456]

Craik FI, Govoni R, Naveh-Benjamin M, Anderson ND. The effects of divided attention on encoding and retrieval processes in human memory. Journal of Experimental Psychology: General. 1996; 125:159–180. [PubMed: 8683192]

Craik FIM, Lockhart RS. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior. 1972; 11:671–684.

Davachi L, Mitchell JP, Wagner AD. Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100:2157–2162. [PubMed: 12578977]

Detre, GJ.; Polyn, SM.; Moore, CD.; Natu, VS.; Singer, BD.; Cohen, JD.; Norman, KA. The Multi-Voxel Pattern Analysis (MVPA) toolbox. Proceedings from Organization for Human Brain Mapping; 2006.

Diana RA, Reder LM. The low-frequency encoding disadvantage: Word frequency affects processing demands. Journal of Experimental Psychology: Learning, Memory and Cognition. 2006; 32:805–815.

Dulas, MR. Doctoral dissertation. Georgia Institute of Technology; GA: 2011. The effect of explicitly directing attention toward item-feature relationships on source memory and aging: an ERP study.

Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science. 1986; 1:54–75.

Ericsson KA, Kintsch W. Long-term working memory. Psychological Review. 1995; 102:211–245. [PubMed: 7740089]

Fanselow MS. Factors governing one-trial contextual conditioning. Learning & Behavior. 1990; 18:264–270.

Fischl B, Salat DH, van der Kouwe AJ, Makris N, Segonne F, Quinn BT, Dale AM. Sequence-independent segmentation of magnetic resonance images. NeuroImage. 2004; 23:S69–84. [PubMed: 15501102]

Gobet F, Lane PC, Croker S, Cheng PC, Jones G, Oliver I, Pine JM. Chunking mechanisms in human learning. Trends in Cognitive Science. 2001; 5:236–243.

Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, Fischl B. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. NeuroImage. 2006; 32:180–194. [PubMed: 16651008]

Hastie, T.; Tibshirani, R.; Friedman, JH. The elements of statistical learning. Springer; 2001.

Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics. 1970:55–67.

Kim K, Yi DJ, Raye CL, Johnson MK. Negative effects of item repetition on source memory. Memory and Cognition. 2012

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. Nature Neuroscience. 2009; 12:535–540.

Lee, H.; Jung, J.; Yi, DJ. Pre-experimental familiarity modulates the effects of item repetition on source memory. Proceedings from Vision Sciences Society; Naples, Florida. 2012.

Lisman JE, Grace AA. The hippocampal-VTA loop: controlling the entry of information into long-term memory. Neuron. 2005; 46:703–713. [PubMed: 15924857]

Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Mazoyer B. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philosophical Transactions of the Royal Society B: Biological Sciences. 2001; 356:1293–1322.

McDuff SG, Frankel HC, Norman KA. Multivoxel pattern analysis reveals increased memory targeting and reduced use of retrieved details during single-agenda source monitoring. Journal of Neuroscience. 2009; 29:508–516. [PubMed: 19144851]

Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review. 1956; 63:81. [PubMed: 13310704]

Moscovitch M. The hippocampus as a “stupid,” domain-specific module: Implications for theories of recent and remote memory, and of imagination. Canadian Journal of Experimental Psychology. 2008; 62:62–79. [PubMed: 18473631]

Moscovitch M, Rosenbaum RS, Gilboa A, Addis DR, Westmacott R, Grady C, Nadel L. Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. Journal of Anatomy. 2005; 207:35–66. [PubMed: 16011544]

Nader K. Memory traces unbound. Trends Neurosci. 2003; 26:65–72. [PubMed: 12536129]

Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends in Cognitive Science. 2006; 10:424–430.

Poppenk J, Kohler S, Moscovitch M. Revisiting the novelty effect: When familiarity, not novelty, enhances memory. Journal of Experimental Psychology: Learning, Memory and Cognition. 2010a; 36:1321–1330.

Poppenk J, McIntosh AR, Craik FIM, Moscovitch M. Past experience modulates the neural mechanisms of episodic memory formation. Journal of Neuroscience. 2010b; 30:4707–4716. [PubMed: 20357121]

Poppenk J, Moscovitch M. A hippocampal marker of recollection memory ability among healthy young adults: contributions of posterior and anterior segments. Neuron. 2011; 72:931–937. [PubMed: 22196329]

Rudy JW, Barrientos RM, O'Reilly RC. Hippocampal formation supports conditioning to memory of a context. Behavioral Neuroscience. 2002; 116:530–538. [PubMed: 12148921]

Rudy JW, O'Reilly RC. Contextual fear conditioning, conjunctive representations, pattern completion, and the hippocampus. Behavioral Neuroscience. 1999; 113:867–880. [PubMed: 10571471]

Rudy JW, O'Reilly RC. Conjunctive representations, the hippocampus, and contextual fear conditioning. Cognitive, Affective and Behavioral Neuroscience. 2001; 1:66–82.

Schacter DL. The seven sins of memory. Insights from psychology and cognitive neuroscience. American Psychologist. 1999; 54:182–203. [PubMed: 10199218]

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Matthews PM. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage. 2004; 23:S208–219. [PubMed: 15501092]

Squire LR, Zola SM. Episodic memory, semantic memory, and amnesia. Hippocampus. 1998; 8:205–211. [PubMed: 9662135]

Stark CE, Okado Y. Making memories without trying: medial temporal lobe activity associated with incidental memory formation during recognition. Journal of Neuroscience. 2003; 23:6748–6753. [PubMed: 12890767]

Tulving, E. Episodic and semantic memory. In: Tulving, E.; Donaldson, W., editors. Organization of memory. Oxford, England: Academic Press; 1972.

Tulving E, Kroll N. Novelty assessment in the brain and long-term memory encoding. Psychonomic Bulletin and Review. 1995; 2:387–390.

Tulving E, Markowitsch HJ, Craik FIM, Habib R, Houle S. Novelty and familiarity activations in PET studies of memory encoding and retrieval. Cerebral Cortex. 1996; 6:71–79. [PubMed: 8670640]

Tulving E, Thomson DM. Encoding specificity and retrieval processes in episodic memory. Psychological Review. 1973; 80:352.

Tulving E, Markowitsch HJ. Episodic and declarative memory: Role of the hippocampus. Hippocampus. 1998; 8:198–204. [PubMed: 9662134]

Wichawut C, Martin E. Independence of AB and AC associations in retroaction. Journal of Verbal Learning and Verbal Behavior. 1971; 10:316–321.

Zou GY. Toward using confidence intervals to compare correlations. Psychological Methods. 2007; 12:399–413. [PubMed: 18179351]

HIGHLIGHTS

1. Classifier read-out of on-task processing predicted later memory for task.
2. Previously known proverbs evoked greater on-task processing than novel or repeated ones.
3. On-task processing was a better memory predictor for repeated proverbs than other types.
4. These effects help explain memory bonuses of pre-study repetition and prior knowledge.
5. Our results suggest a dissociation in the mechanisms underlying such memory bonuses.

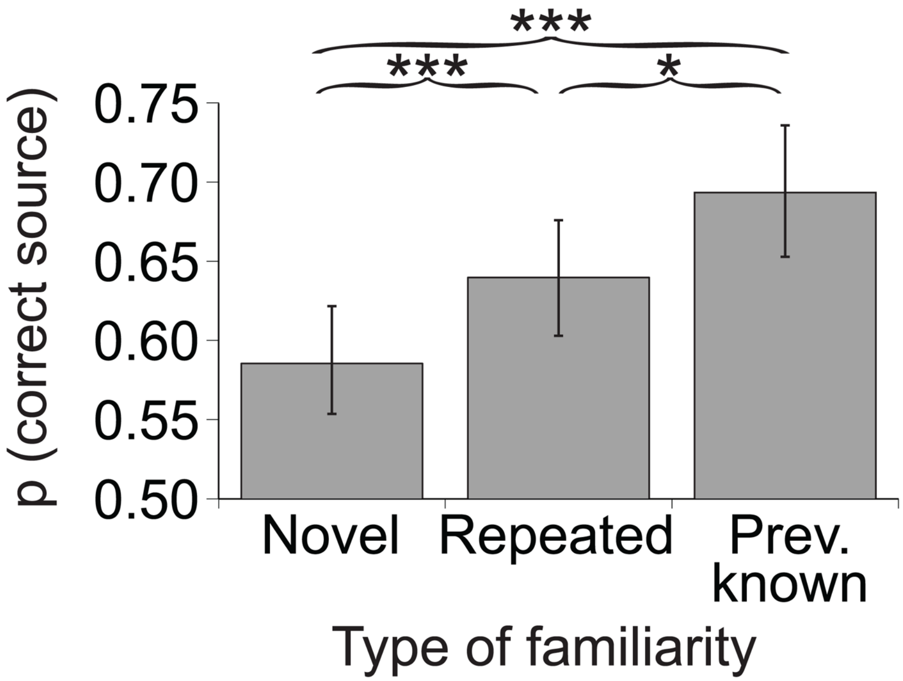

Fig. 1. Source memory performance

The proportion of correct source responses was higher for repeated and previously known proverbs than for novel proverbs. The symbols *** and * designate statistically significant differences at $P < 0.001$ and at $P < 0.05$, respectively.

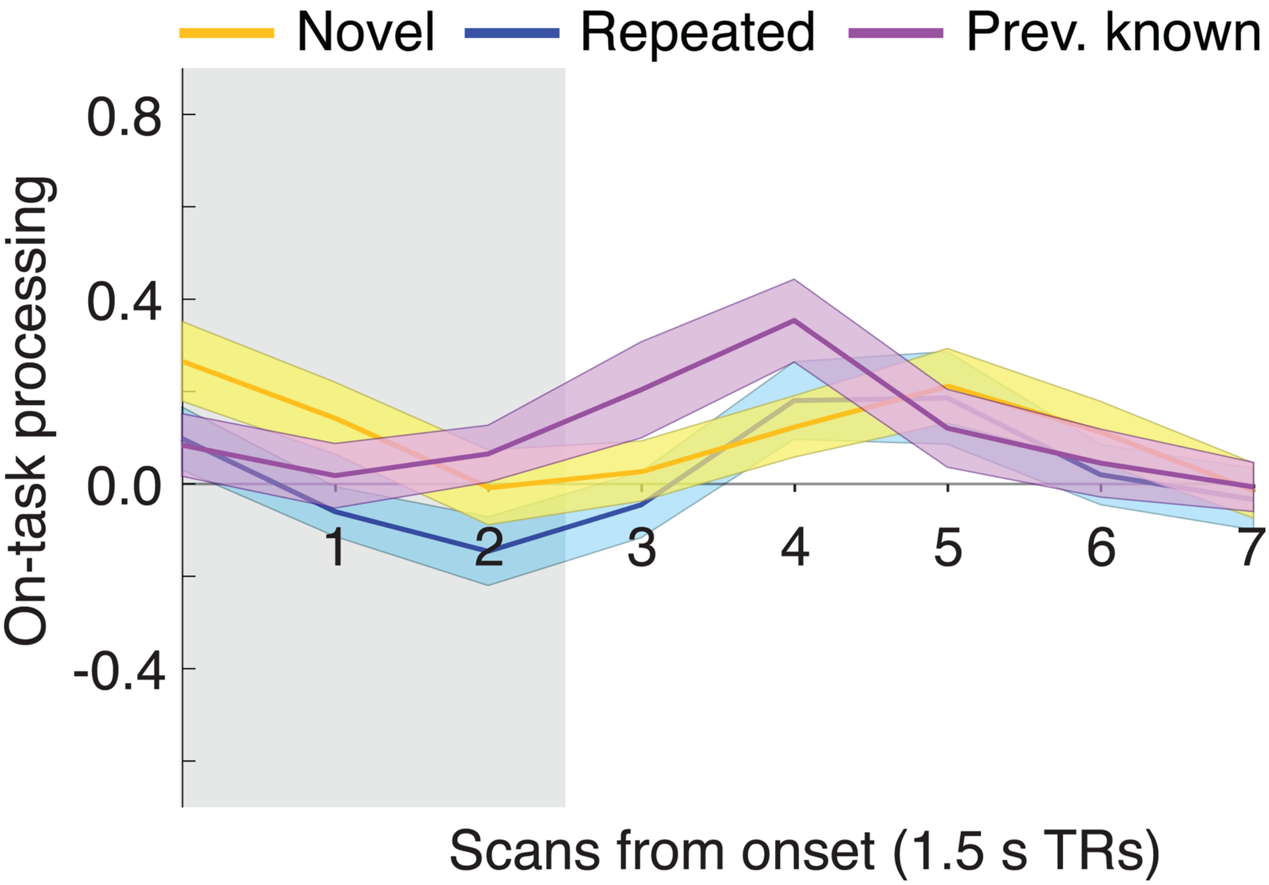

Fig. 2. On-task processing within novelty condition

Line plots show mean Phase 2 on-task processing for novel, repeated, and previously known proverbs, for the eight scans that were acquired post-stimulus onset (each scan was collected over a period of 1.5 s). Ribbons about the mean designate ±1 bootstrap standard error. Grey shading indicates the period during which the stimulus was on screen (no hemodynamic shift was applied to the response function). See also Fig. S1 for a model optimization parameter search and Fig. S2 for on-task processing importance maps.

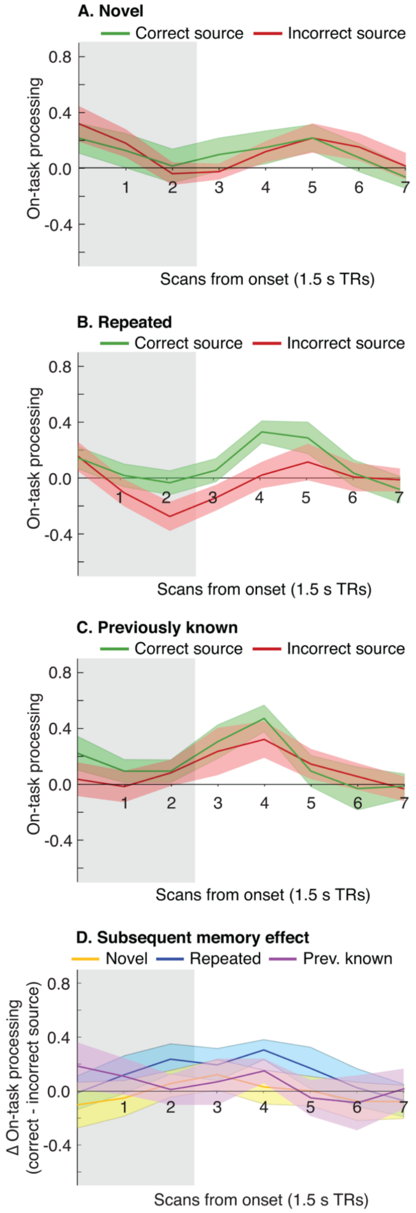

Fig. 3. On-task processing as memory predictor

Line plots show mean levels of Phase 2 on-task processing for the eight scans that were acquired post-stimulus onset, plotted separately as a function of whether the source was subsequently remembered correctly in Phase 3. These plots were made separately for novel, repeated, and previously known proverbs (A–C). Panel (D) shows the subsequent memory effect (the difference in Phase 2 on-task processing for items where the source was subsequently remembered correctly vs. incorrectly) for each of the novelty conditions. Ribbons about the mean designate ±1 bootstrap standard error. Grey shading indicates the period during which the stimulus was on screen (no hemodynamic shift was applied to the response function).

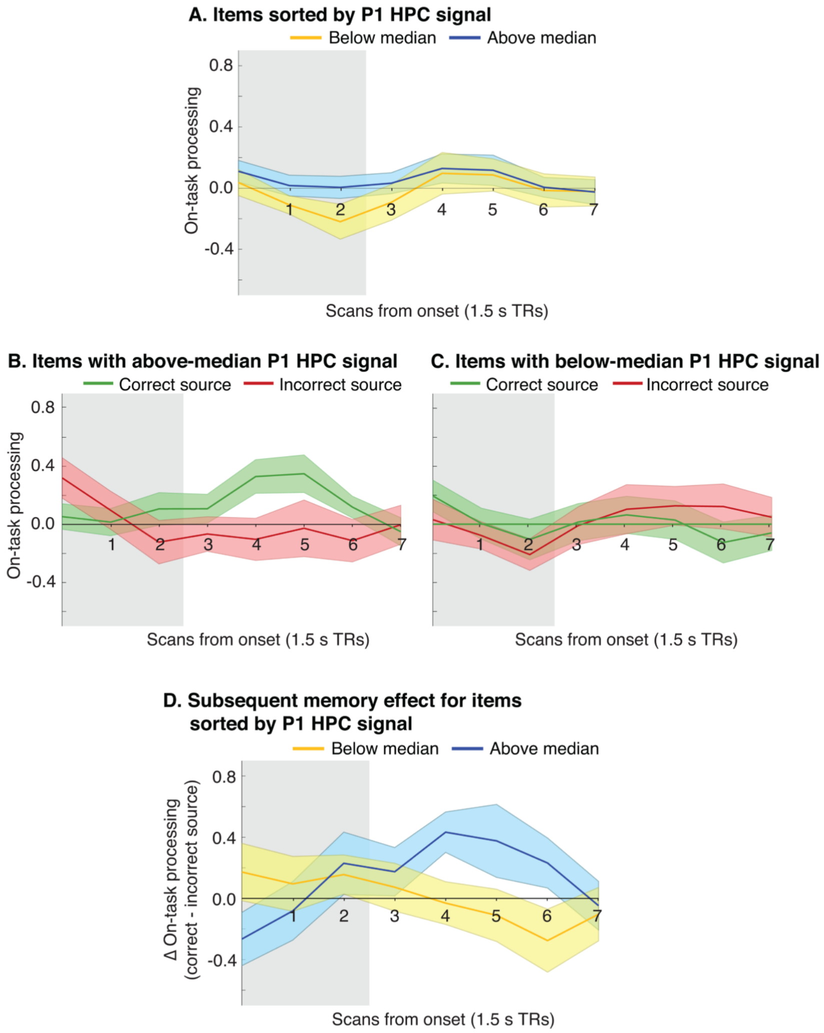

Fig. 4. Sorting of repeated items by Phase 1 hippocampal signal

Repeated items were split into two groups based on whether the Phase 1 hippocampal (HPC) signal was above or below the median. Plot (A) shows mean levels of Phase 2 on-task processing for each of the two groups (above-median and below-median) for the eight scans that were acquired beginning at stimulus onset. Plots (B) and (C) show mean levels of Phase 2 on-task processing, further split by whether the item's source was subsequently remembered correctly in Phase 3. Plot (D) shows the subsequent memory effect (the difference in Phase 2 on-task processing for items where the source was subsequently remembered correctly vs. incorrectly) for above-median and below-median items. Ribbons about the mean designate ±1 bootstrap standard error. Grey shading indicates the period during which the stimulus was on screen (no hemodynamic shift was applied to the response function).

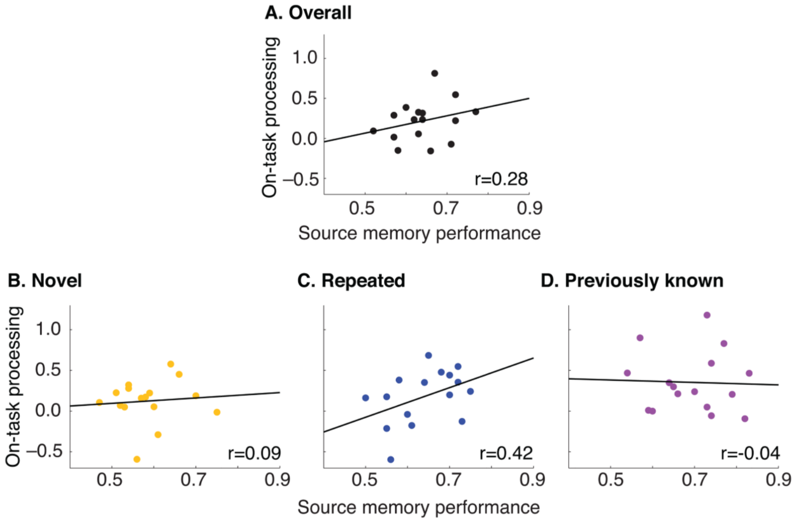

Fig. 5. On-task processing and memory across participants

Scatterplots show each participant's performance on the Phase 3 source memory task (i.e., percent correct source responses) as a function of their average Phase 2 on-task processing classifier score, with linear fits. This relationship is plotted when the data are collapsed across conditions (A) and also separately for novel, repeated, and previously known proverbs (B–D).

Table 1

Schematic of experimental protocol and stimulus exposure. Stimuli consisted of two lists of 40 English proverbs (English$_1$ and English$_2$) and four lists of 20 Asian proverbs (Asian-Repeated$_1$, Asian-Repeated$_2$, Asian-Novel$_1$, and Asian-Novel$_2$).

Phase and purpose	**Lists presented and task instructions**	
Phase I. Three repetitions for familiarity induction	Asian-Repeated$_1$	Asian-Repeated$_2$
	multiple tasks	
Phase II. Incidental encoding of proverbs in two tasks	English$_1$ Asian-Repeated$_1$	English$_2$ Asian-Repeated$_2$
	Asian-Novel$_1$	Asian-Novel$_2$
	rate quality	*rate target age*
Phase III. Test of memory for Phase II source information	(all items)	
	rated quality or target age?	
Follow-up. Identification of proverbs known prior to the experiment	(all items)	
	learned today, or previously known?	

Table 2

Reliability of task classification, as a function of novelty condition and time point

For each time point and novelty condition, did the classifier output for the correct task reliably exceed the classifier output for the incorrect task? Each cell contains the bootstrap ratio associated with the comparison.

	TR							
Condition	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Novel	3.09*	1.81*	−0.09	0.42	1.70*	2.47*	1.56	−0.32
Repeated	1.40	−1.12	−2.15	−0.76	2.33*	1.94*	0.42	−0.41
Previously known	1.15	0.11	1.09	1.96*	4.07*	1.61	0.79	0.45

*$P < 0.05$, one-tailed

Table 3

Reliability of differences between novelty conditions, as a function of time point

Left side: Were levels of Phase 2 on-task processing different across novelty conditions? Right side: Was the size of the Phase 2 subsequent memory effect (the difference in Phase 2 on-task processing, as a function whether of the item's source was subsequently remembered correctly vs. incorrectly during Phase 3) different across novelty conditions? Each cell contains the bootstrap ratio associated with the comparison.

Time (TR)	Contrast					
	Phase 2 on-task processing			Phase 2 on-task processing subsequent memory effect		
	Repeated > Novel	Previously known > Novel	Previously known > Repeated	Repeated > Novel	Previously known > Novel	Previously known > Repeated
3	−0.80	−0.84	3.59*	0.70	−0.29	−0.65
4	0.71	2.41*	2.00*	2.40*	0.92	−1.60
5	−0.20	−0.98	−0.71	0.90	−0.27	−0.99

*$P < 0.05$

Table 4

Reliability of effects of Phase 1 encoding strength, as a function of time point

Left side: Were levels of Phase 2 on-task processing different for items where Phase 1 hippocampal (HPC) activity was above the median vs. below the median? Right side: Was the size of the Phase 2 subsequent memory effect (the difference in Phase 2 on-task processing, as a function of whether the item's source was subsequently remembered correctly vs. incorrectly during Phase 3) different for items where Phase 1 HPC activity was above the median vs. below the median? Each cell contains the bootstrap ratio associated with the comparison.

Time (TR)	Above > Below-median Phase 1 HPC signal sorting of Phase 2 repeated items	
	Phase 2 on-task processing	Phase 2 on-task processing subsequent memory effect
3	0.44	0.95
4	0.45	2.70*
5	0.88	1.46

* $P < 0.05$