Resistance to forgetting associated with hippocampus-mediated reactivation during new learning

Brice A Kuhl[1], Arpeet T Shah[1], Sarah DuBrow[1] & Anthony D Wagner[1,2]

One of the reasons why we forget past experiences is because we acquire new memories in the interim. Although the hippocampus is thought to be important for acquiring and retaining memories, there is little evidence linking neural operations during new learning to the forgetting (or remembering) of earlier events. We found that, during the encoding of new memories, responses in the human hippocampus are predictive of the retention of memories for previously experienced, overlapping events. This brain-behavior relationship is evident in neural responses to individual events and in differences across individuals. We found that the hippocampus accomplishes this function by reactivating older memories as new memories are formed; in this case, reactivating neural responses that represented monetary rewards associated with older memories. These data reveal a fundamental mechanism by which the hippocampus tempers the forgetting of older memories as newer memories are acquired.

Although it is well established that successful encoding of new episodic memories depends on the hippocampus[1,2], successful encoding alone does not guarantee long-lasting retention. Instead, a variety of factors determine whether memories will ultimately be remembered or forgotten after encoding occurs[3–6]. The risk of forgetting is particularly high when initial encoding events are followed by similar or overlapping experiences, creating interference between the past and present[7–9]. A primary challenge for theories of hippocampal function and episodic memory is to understand how new learning is balanced against the forgetting of past memories.

Computational models of hippocampal function emphasize two core mechanisms that are thought to guard against forgetting: pattern separation and pattern completion[10,11]. Pattern separation refers to the orthogonal coding of memories for overlapping events, which can reduce forgetting by creating distinct (non-interfering) representations[10,12–14]. Pattern completion, on the other hand, allows previously encoded memories to be reinstated from a partial input[15], thereby allowing past episodes to be reactivated, interleaved with current experience, and consolidated over time[16,17]. To date, there is a marked lack of evidence directly linking the engagement of either of these mechanisms during new learning to specific instances of remembering or forgetting of past events.

At the same time, there is a growing body of evidence from electrophysiological studies of rodents indicating that patterns of neural activity associated with past events can be reactivated via coordinated interactions between the hippocampus and other neocortical or subcortical structures[18–20]. Although such reactivation has most typically been observed in the form of spontaneous firing patterns during sleep, recent evidence indicates that reactivation also occurs during awake behavior[21,22] and can be triggered by external cues in humans[23], suggesting a potential role for reactivation during ongoing learning. Although reactivation represents a form of pattern completion that is widely hypothesized to support the consolidation of memories[24], the critical link between hippocampus-mediated reactivation of individual memories and diminished forgetting of those memories has yet to be demonstrated.

We used functional magnetic resonance imaging (fMRI) to track neural responses during the encoding of overlapping events to determine whether hippocampal operations during the ongoing encoding of new events actively limit the forgetting of previously acquired memories of past events. Moreover, to the extent that forgetting of the past is diminished as a result of hippocampal processes during subsequent learning, we further sought to determine whether this relationship reflects the benefits of pattern completion, whereby memories of past events are reactivated during new learning, or pattern separation, whereby new memories are represented as being distinct from memories of past events.
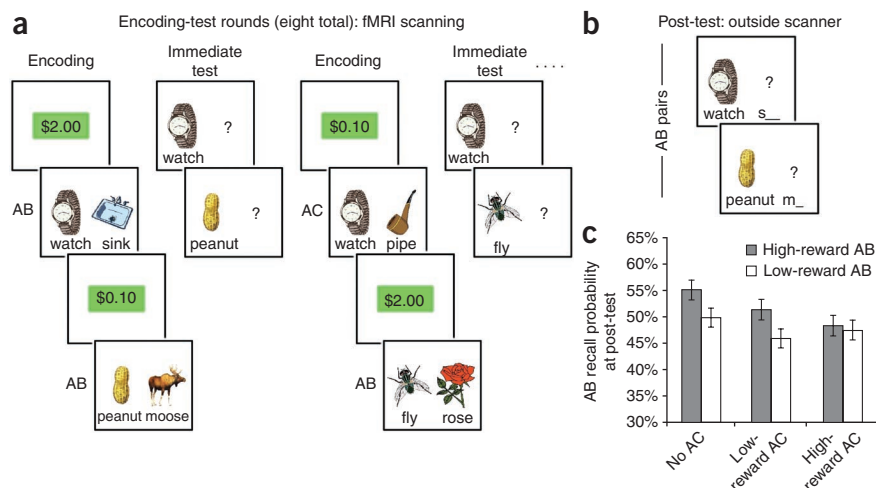
To assess the relationship between the encoding of new memories and the retention of older memories, we adapted an AB-AC learning protocol for use with human subjects during fMRI. During scanning, subjects alternated between periods of encoding and retrieval. During encoding, subjects studied pairs of items presented to the left (cue) and right (associate) of center. The pairs consisted of either novel cues with novel associates (AB) or repeated cues with novel associates (AC) (**Fig. 1a** and **Supplementary Fig. 1**). During retrieval, each cue from the preceding encoding phase was presented and subjects attempted to retrieve the relevant associate. Each AB pair was encoded and retrieved before a corresponding AC pair was encoded and retrieved. To motivate learning of AB and subsequent AC events, each pair was associated with a potential monetary reward. The level of reward was manipulated (high versus low) to provide a contextual element associated with each pair that would differentially elicit engagement of

[1]Department of Psychology and [2]Neurosciences Program, Stanford University, Stanford, California, USA. Correspondence should be addressed to B.A.K. (brice.kuhl@yale.edu) or A.D.W. (awagner@stanford.edu).

Figure 1 Experimental design and behavioral results. (**a**) During the encoding rounds, subjects studied pairs of items. Pairs consisted of either a novel cue paired with a novel associate (AB pair; for example, 'watch-sink') or a repeated cue paired with a novel associate (AC pair; for example, 'watch-pipe'). Two-thirds of all AB pairs were followed by a corresponding AC pair in the subsequent encoding round; the remaining one-third of AB pairs were not associated with corresponding AC pairs. Trials began with the presentation of a reward value ($2.00 or $0.10, high reward or low reward, respectively), indicating the potential value for later remembering the upcoming pair (see Online Methods and **Supplementary Fig. 1**). Thus, every AB pair was associated with either high or low reward and was later followed by a corresponding high reward AC pair, a low reward AC pair or no AC pair. Each encoding round was followed by an 'immediate test' round, during which subjects were shown each cue (A terms) from the immediately preceding encoding round and attempted to recall the corresponding associate. In this manner, each AB pair was encoded and tested before the corresponding AC pair was encoded. (**b**) After eight alternating rounds of encoding and immediate test, a critical post-test was administered outside of the scanner, during which subjects were cued to recall each previously encoded AB pair, both those that had been followed by corresponding AC pairs (interference condition) and those that were not followed by corresponding AC pairs (no interference condition). (**c**) Performance on the post-test revealed that AB pairs followed by AC pairs were more likely to be forgotten, reflecting the deleterious effect of retroactive interference. Error bars indicate ± within-subject error.

reward-related neural mechanisms. Half of all of the AB pairs were associated with high reward and half with low reward; likewise for AC pairs (AB reward level was independent of AC reward level). After the encoding and retrieval phases, a separate reward-anticipation task with no mnemonic component was administered to independently localize reward-sensitive regions[25]. Subjects then exited the scanner and completed a critical post-test that probed subsequent memory for all AB pairs (**Fig. 1b**).

RESULTS
Behavioral performance
To assess the effect that new learning (AC pairs) has on memory of older, overlapping events (AB pairs), we evaluated memory performance on the post-test. Replicating classic retroactive interference effects[7,8], AB pairs that were followed by overlapping AC pairs were more poorly remembered at post-test, relative to AB pairs not followed by an overlapping AC pair ($P < 0.05$, ANOVA, $n = 20$; **Fig. 1c** and **Supplementary Table 1**). Subjects varied considerably in the amount of interference-induced forgetting that they suffered from, allowing for consideration of neural factors that relate to individual differences in forgetting.

Consistent with prior evidence that reward anticipation benefits declarative memory[26,27], there was a modest trend toward better memory for AB pairs associated with high, relative to low, reward ($P = 0.11$; **Fig. 1c**), an effect that was significant ($P < 0.05$) in a separate behavioral experiment (**Supplementary Results** and **Supplementary Table 1**). AB reward level (high versus low reward) did not interact with AC reward level (no AC, low reward, high reward) ($P = 0.36$; **Fig. 1c**), indicating that high and low reward AB pairs were similarly affected by interference (for additional behavioral results, see **Supplementary Tables 2–5**).

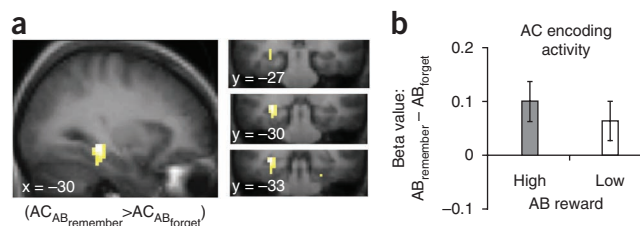### Neural AC encoding responses and subsequent AB memory
The primary goal of the fMRI experiment was to determine how neural mechanisms engaged during the encoding of new events related to subsequent memory of previously encoded events. Accordingly, we evaluated fMRI data acquired during the encoding of AC pairs,

separating trials as a function of later memory (that is, remembered versus forgotten) for the previously encoded, corresponding AB pairs (**Supplementary Table 6**). This trial-level subsequent memory analysis[28,29] revealed that, during AC encoding, greater activation in the left posterior hippocampus and parahippocampal cortex was associated with better memory (that is, reduced forgetting) of corresponding AB pairs (**Fig. 2** and **Supplementary Table 7**). This observation suggests that processes subserved by the hippocampus limit forgetting of previously acquired, overlapping memories during the encoding of new memories (for related results, see **Supplementary Table 8**).

Building on the finding that trial-by-trial differences in hippocampal activation during AC encoding are related to resistance to AB forgetting, we next determined whether a similar relationship is evident across individuals. That is, do differences in hippocampal activation during AC encoding differentiate individuals on the basis of their susceptibility to interference-related forgetting of AB pairs? As noted above, individuals varied widely in their susceptibility to interference-related forgetting. We conducted a whole-brain regression analysis using the contrast of all AC encoding trials versus baseline regressed against each subject's interference-related forgetting score (retroactive interference). A negative correlation was observed in left posterior hippocampus and parahippocampal cortex (**Fig. 3a,b** and **Supplementary Table 9**); that is, greater activation was associated with less retroactive interference–induced forgetting, indicating that individual differences in hippocampal engagement during new learning are related to susceptibility to retroactive interference. Notably, conjunction analysis confirmed that the localization of this across-subject effect was anatomically convergent with the independent within-subject, trial-by-trial effect (**Fig. 3c**). The convergence of these two analyses suggests a relationship between hippocampal activity during AC encoding and resistance to AB forgetting.

Although these analyses suggest that mechanisms engaged during AC encoding relate to AB retention, it is important to rule out two alternative explanations. First, hippocampal activation during AC encoding may simply reflect the extent to which AB pairs were initially learned, which in turn relates to AB memory at post-test. Enabled by the fact that memory for AB pairs was tested both prior

Figure 2 Relationship between AC encoding and AB forgetting.
(**a**) Activation in the posterior hippocampus (Montreal Neurological Institute coordinates: –30, –33, –9) extending into parahippocampal cortex (–30, –30, –18) during AC encoding was associated with spared forgetting of AB pairs ($P_{FDR} < 0.05$ for both regions, small volume correction at voxel level; for complete results, see **Supplementary Table 7**). (**b**) Beta values showing the relationship between AC encoding activation and subsequent AB memory separately for AB pairs associated with high and low reward; drawn from the hippocampal ROI from contrast depicted in **a**. Error bars indicate ± within-subject error. FDR, false discovery rate. Small volume correction was conducted using Anatomical Automatic Labeling atlas (http://www.cyceron.fr/web/aal_anatomical_automatic_labeling.html) to generate mask of entire medial temporal lobe, including hippocampus, parahippocampal cortex, perirhinal cortex and entorhinal cortex.

and subsequent to AC encoding, we conducted analyses to explore this possibility. Using the hippocampal region of interest (ROI) identified from the within-subject analysis, we carried out a multiple regression analysis and found that, although between-subject differences in hippocampal activation during AC encoding were negatively related to AB forgetting ($P < 0.05$, $n = 19$), this hippocampal effect was unrelated to between-subject differences in initial AB learning ($P > 0.5$). Similarly, the ROI identified from the between-subject analysis was submitted to an independent within-subject analysis using a new general linear model ($n = 19$; **Supplementary Table 10**) in which AC encoding activation was considered as a function of both initial AB learning (before AC encoding) and later AB retention (following AC encoding). Notably, hippocampal activation during AC encoding was greater when corresponding AB pairs were initially learned and subsequently retained, relative to initially learned and subsequently forgotten ($P = 0.06$) or both initially and subsequently forgotten ($P < 0.05$); activation for these latter two cases did not differ ($P > 0.8$) (**Fig. 3d**). Finally, a separate voxel-level analysis, based on this second model and restricted to AC encoding events for which the corresponding AB events were initially learned, revealed that activation in the hippocampus during AC encoding was positively associated with retaining the corresponding AB pair (at a slightly relaxed threshold, $P < 0.005$; **Fig. 3e** and **Supplementary Table 11**). Collectively, these analyses confirm that there

is a relationship between hippocampal responses during AC encoding and AB retention, even when controlling for AB learning.

A second potential interpretation of our hippocampal findings is that, if memory for AB and AC pairs is highly correlated (that is, when the B term is learned, the C term is more likely to be learned), then the relationship between AC encoding and AB retention might simply reflect an underlying relationship between AC encoding and subsequent AC recall. Although behavioral evidence indicated a modest positive relationship between AB and AC learning (**Supplementary Results** and **Supplementary Tables 4** and **5**), directed fMRI analyses revealed that the relationship between hippocampal activation during AC encoding and subsequent AB memory was independent of AC learning. Specifically, although hippocampal activation during AC encoding was positively related to both AB retention and AC learning, these effects did not interact (**Supplementary Tables 12** and **13**, **Supplementary Results** and **Supplementary Figs. 2** and **3**).

Reward-related AB reactivation during AC encoding
We next sought to determine whether resistance to interference was established via pattern completion, whereby AB pairs were reactivated during AC encoding. On their own, the hippocampal data are consistent with either a pattern separation or pattern completion mechanism. That is, pattern separation may have facilitated AB retention during

Figure 3 Hippocampal responses during encoding and susceptibility to retroactive interference. (**a**) Between-subject regression of AC encoding activation against proportionalized retroactive interference (RI) revealed a negative relationship between activation in hippocampus (–36, –30, –12), parahippocampal cortex (–30, –36, –15) and the magnitude of retroactive interference ($P_{FDR} \leq 0.05$ for both regions, small volume correction at voxel level; for complete results, see **Supplementary Table 9**). (**b**) Scatter plot showing activation in hippocampus as function of retroactive interference; higher activation during AC encoding was associated with reduced retroactive interference. (**c**) Conjunction of between-subject regression analysis (described in **a** and **b**; shown in red) and within-subject analysis (described in **Fig. 2a,b**; shown in yellow) revealed overlap (orange) in the hippocampus and parahippocampal cortex (for display purposes only, each contrast is thresholded at $P < 0.005$, uncorrected). (**d**) Hippocampal activation during AC encoding (extracted from the region observed in the between-subject regression analysis) as a function of initial learning (Pre-AC) and later memory (Post-AC) revealed a selective increase in hippocampal response during AC encoding when AB pairs were initially learned and subsequently retained. (**e**) Voxel-level contrast of AC encoding trials associated with initial AB learning and subsequent AB retention versus initial AB learning and subsequent AB forgetting revealed activation in hippocampus (27, –30, –6; $P < 0.005$, uncorrected). Error bars indicate ± within-subject error.
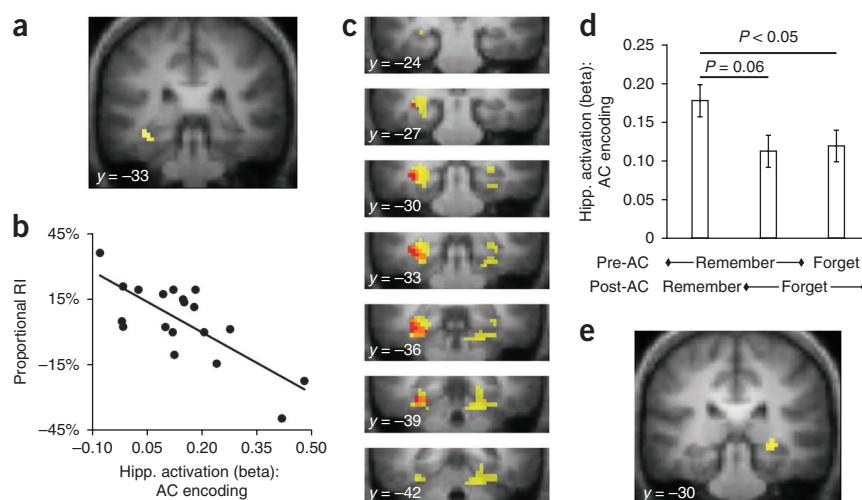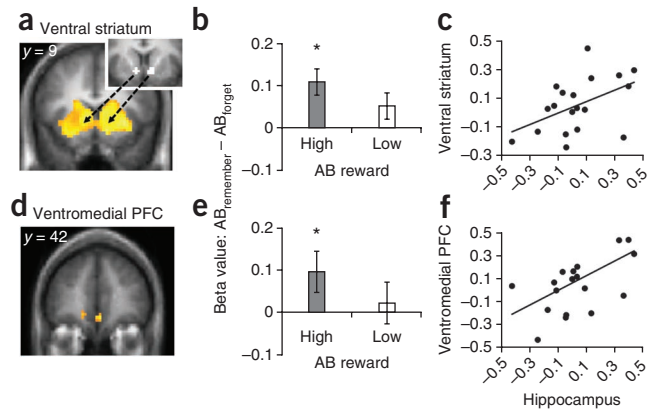
Figure 4 ROI analysis of reward-sensitive regions, as defined from independent reward-localizer task. (**a**) Contrast of high-reward anticipation versus low-reward anticipation from the reward-localizer task revealed activation in dorsal and ventral striatum ($P < 0.001$, uncorrected). Inset shows anatomical mask applied to functional data to obtain ventral striatum ROI, which was then applied to the encoding data (**Supplementary Results**). (**b**) Activation in ventral striatum during AC encoding predicted subsequent memory for AB pairs that were associated with high reward (*$P < 0.05$). (**c**) Across subjects, a relationship was observed between AC encoding responses in ventral striatum and hippocampus. Specifically, the greater the bias in ventral striatum toward predicting subsequent memory for high versus low reward AB pairs ((High $AB_{remember}$ − High AB_{forget}) − (Low $AB_{remember}$ − Low AB_{forget})), the greater the bias in hippocampus (correlation coefficient $r = 0.473$, $P < 0.05$). (**d**) Contrast of hits versus misses from the reward-localizer task (for details see **Supplementary Results**) revealed activation in vmPFC ($P < 0.005$, uncorrected). All vmPFC voxels that showed this effect were combined into a single region of interest that was then applied to the encoding data (see **Supplementary Results**). (**e**) Data are presented as in **b** but for activation in vmPFC (*$P < 0.05$). (**f**) Data are presented as in **c** but for the relationship between activation in vmPFC and hippocampus ($r = 0.478$, $P < 0.05$). Error bars indicate ± within-subject error.

AC encoding by allowing AC pairs to be neurally represented as being orthogonal to AB pairs, thereby reducing interference and forgetting. Indeed, recent fMRI data indicate that hippocampal activation increases during conditions in which pattern separation is likely to occur[12]. However, although pattern separation would serve to reduce interference and promote AB retention, pattern separation inherently does not involve the reactivation of AB representations during AC encoding[10]. Thus, a pattern completion account of our hippocampal data uniquely predicts that the relationship between hippocampal activation during AC encoding and resistance to AB forgetting reflects reactivation of the previously encoded AB events, in direct response to AC pairs, thereby promoting AB retention.

To assess the contribution of pattern completion, we examined whether contextual elements associated with AB pairs were engaged during AC encoding and in relation to later AB memory. Specifically, as all AB pairs were associated with monetary rewards, we probed responses in two regions that have repeatedly been implicated in processing reward: ventromedial prefrontal cortex (vmPFC) and ventral striatum[30,31]. In particular, ventral striatum has been associated with facilitating reward-motivated declarative memory formation[26,27] and recent evidence suggests that spontaneous reactivation of reward-related memories is associated with coupled responses in ventral striatum and hippocampus in rodents[32–34]. Accordingly, if the hippocampus limits AB forgetting during AC encoding by reactivating AB pairs, this reactivation should be evident in ventral striatum and vmPFC regions that represent AB reward associations. Specifically, retention of AB pairs should be associated with heightened activation of reward-related regions during AC encoding.

Using data from the independent reward localizer task, we generated ventral striatum and vmPFC ROIs consisting of voxels in each region that were sensitive to reward values in the localizer task (**Fig. 4** and **Supplementary Results**). During AB encoding, activation in these regions was greater for high versus low reward pairs (ventral striatum, $P = 0.05$; vmPFC, $P < 0.01$), confirming their sensitivity to AB reward context. During AC encoding, however, AC reward values did not modulate activation in these regions (P's > 0.9), suggesting that the mnemonic history, namely past rewards, associated with AC pairs influenced reward-related responses during present encoding. Notably, as predicted by the pattern completion (reactivation) hypothesis, during AC encoding, activation in these regions, which were selected only on the basis of their sensitivity to reward in an independent task, was positively associated with AB retention. That is, greater activation in these regions during AC encoding was associated with better memory for previously encoded AB pairs ($P < 0.05$). Follow-up analyses indicated that this relationship was significant in each ROI for AB pairs associated with high reward (P's < 0.05), but not for AB pairs associated with low reward (ventral striatum, $P = 0.15$; vmPFC, $P = 0.67$) (**Fig. 4b,e**). However, the interaction between AB reward level and subsequent memory was not significant ($P = 0.23$; for additional consideration of how these data and the hippocampal data relate to reward values, see **Supplementary Figs. 4** and **5**).

To further test the pattern completion account, we assessed whether there was a relationship between activation in the hippocampus and reward-related regions during AC encoding. Across subjects, the magnitude of the hippocampal subsequent memory effect was highly correlated with the magnitude of the subsequent memory effects in ventral striatum (correlation coefficient $r = 0.70$, $P < 0.005$) and vmPFC ($r = 0.75$, $P < 0.001$). Moreover, the difference in the magnitude of the hippocampal subsequent memory effect for high- versus low-reward AB pairs was correlated with this same difference in reward-related regions (ventral striatum, $r = 0.47$, $P < 0.05$; vmPFC, $r = 0.48$, $P < 0.05$; **Fig. 4c,f**). These correlations are consistent with a pattern completion account of the hippocampal data, in which hippocampal responses drive reactivation of frontostriatal regions. Together, these findings suggest that, during AC encoding, hippocampal pattern completion processes reactivated previously encoded AB pairs, along with their associated reward context, and that this reactivation protected AB memories against interference-based forgetting.

DISCUSSION

Theoretical understanding of the functional neurobiology of event memory requires specification of the mechanisms that enable learning while mitigating forgetting. Our results provide evidence implicating the hippocampus in minimizing the forgetting that is brought about by new learning, arguably the most ubiquitous form of forgetting. In particular, hippocampal activation during new learning was predictive of which past memories would be most resistant to forgetting, as well as which individuals would be less susceptible to forgetting. Moreover, frontostriatal regions that represented reward values associated with past events were also activated during subsequent overlapping events to the extent that the older memories were retained. Across subjects, these frontostriatal responses were correlated with hippocampal responses, consistent with the idea that hippocampal pattern completion drives frontostriatal reactivation[32–34] and that this reactivation supports the retention of past memories.

An important tenet of computational theories of hippocampal function is that the hippocampus regulates the balance between memories of past and present events during new learning[10,17]. Although this balance is thought to be achieved via both pattern separation and pattern completion, our results are most readily accounted for in terms of pattern completion. Specifically, a pattern completion account predicts that AB events and their associated neural representations should be reactivated during AC encoding, on account of the shared A term, thereby promoting AB retention over the long term. The pattern of activation observed in ventral striatum and vmPFC during AC encoding was highly consistent with this prediction. In particular, marked engagement of ventral striatum and vmPFC was observed during AC encoding when previously encoded AB events were associated with high rewards and later remembered. Although this pattern of responses in frontostriatal regions suggests that pattern completion is involved in protecting AB memories against forgetting, it is important to note that these data do not preclude the possibility that pattern separation contributed to our results. For example, although some subregions of the hippocampus may have been biased toward pattern completion, other subregions may have been biased toward pattern separation[12]. In addition, individual trials, or even time points in a trial, may have varied in the degree to which they elicited pattern completion versus separation. Thus, although it is not clear that pattern separation would produce the observed relationship between frontostriatal activation during AC encoding and AB retention, and pattern separation does not predict the observed independence between the subsequent memory effects for AB versus AC pairs measured during AC encoding, it is nevertheless possible that the observed hippocampal responses reflect a blend of pattern completion and pattern separation mechanisms.

Although prior evidence of neural reactivation has often been recorded during periods of sleep[18–20,24,35,36], computational theories of hippocampal function have emphasized that reactivation may occur whenever cues associated with past events are re-encountered, thereby eliciting pattern completion[17]. Indeed, there have been several reports of awake reactivation[21,22,37], including recent evidence that awake reactivation is of higher fidelity than reactivation in more sleep-like states[22]. Although these data suggest that awake reactivation may be particularly well suited to promoting memory retention[22], it has been alternatively hypothesized that reactivation is more likely to promote the durability of reactivated traces if it occurs during sleep rather than during awake behavior[24]. Specifically, during awake behavior, external inputs are posited to increase the risk that reactivated traces will be destabilized or disrupted during reconsolidation[24,38] and it has been speculated that retroactive interference may cause forgetting precisely because older memories are reactivated during new learning[5,6]. From this alternative perspective, our results are surprising, as they indicate that trace reactivation in the presence of ongoing external input (that is, AC pairs) confers mnemonic benefits.

One way in which reactivation may have strengthened older memories in spite of ongoing learning is via integration. Specifically, during AC encoding, reactivated B terms may have been directly integrated into encoded representations of AC pairs. By this account, reactivation alone may not fully account for our results; instead, reactivation may have enabled integration to occur, which conferred critical benefits to reactivated memories. Consistent with this notion, prior behavioral evidence indicates that integration across potentially interfering memories can markedly reduce forgetting[39]. Similarly, recent evidence suggests that existing knowledge structures allow new, related information to be more readily consolidated[40]. Notably, these forms of integration are likely supported by the hippocampus, as the hippocampus has been shown to support integration across events that share feature overlap[41–43]. In particular, recent evidence suggests that the hippocampus supports integration during ongoing learning, thus enabling associative inference[42,44]. Although computational theories of memory have emphasized integration as a means of reducing forgetting across overlapping events[45], these models postulate a form of integration in which event-specific details are lost in favor of more generalized statistical learning. In contrast, our results would require a form of integration that preserves the details that differentiate past and present events (that is, B and C event features).

As an alternative to integration, it is possible that reactivation of AB pairs allowed for a form of distinctive encoding in which subjects used reactivated B terms to help orient them to non-overlapping features of C terms. Mechanistically, this account is distinct from pattern separation in that it can only occur to the extent that B terms are reactivated (pattern completed), but it shares a conceptual similarity in that it suggests a means by which the distinction between B and C terms could be maintained (and confusability minimized).

Our results build on accumulating evidence concerning the role of ventral striatum and vmPFC in reward-related learning. Ventral striatum, in particular, has been repeatedly implicated in processing rewards in mnemonic[26,27] and non-mnemonic contexts[25,31] and has recently been shown to exhibit reactivation following reward-associated learning tasks[32–34], potentially representing reactivation of value information associated with individual memories[32]. Our results strongly converge with these findings, but further indicate that representations of rewards associated with past events are reactivated in direct response to current environmental cues. This suggests a means by which the motivational importance of past events can be incorporated into current learning experiences, a process that would likely confer adaptive benefits.

Our findings underscore the dynamic nature of episodic memory, wherein the fate of individual memories can be markedly influenced by subsequent mnemonic activities[3–6]. Our data suggest a link between hippocampus-mediated reactivation of individual memories during ongoing learning and the ultimate retention of these reactivated memories. More broadly, hippocampus-mediated reactivation of past events in relation to present experience may support the flexible integration of discrete learning experiences[41–43] and may also serve to reduce the distortions of memory that often occur as a result of experiences that follow initial learning events[46]. In this manner, the hippocampus eases the conflict between older and newer memories, allowing us to learn in the present while retaining the past.

METHODS
Methods and any associated references are available in the online version of the paper at http://www.nature.com/natureneuroscience/.

Note: Supplementary information is available on the Nature Neuroscience website.

AUTHOR CONTRIBUTIONS
B.A.K. and A.D.W. designed the experiments and prepared the manuscript. B.A.K., A.T.S. and S.D. contributed to data collection and analysis.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Squire, L.R., Stark, C.E. & Clark, R.E. The medial temporal lobe. *Annu. Rev. Neurosci.* **27**, 279–306 (2004).
2. Eichenbaum, H. Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron* **44**, 109–120 (2004).
3. Dudai, Y. The neurobiology of consolidations, or, how stable is the engram? *Annu. Rev. Psychol.* **55**, 51–86 (2004).
4. Frankland, P.W. & Bontempi, B. The organization of recent and remote memories. *Nat. Rev. Neurosci.* **6**, 119–130 (2005).
5. Hupbach, A., Gomez, R., Hardt, O. & Nadel, L. Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. *Learn. Mem.* **14**, 47–53 (2007).
6. Hupbach, A., Hardt, O., Gomez, R. & Nadel, L. The dynamics of memory: context-dependent updating. *Learn. Mem.* **15**, 574–579 (2008).
7. Müller, G.E. & Pilzecker, A. Experimentelle Beiträge zur Lehre vom Gedächtnis. *Z. Psychol. Ergänzungsband* **1**, 1–300 (1900).
8. Barnes, J.M. & Underwood, B.J. Fate of first-list associations in transfer theory. *J. Exp. Psychol.* **58**, 97–105 (1959).
9. Anderson, M.C. Rethinking interference theory: executive control and the mechanisms of forgetting. *J. Mem. Lang.* **49**, 415–445 (2003).
10. O'Reilly, R.C. & McClelland, J.L. Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. *Hippocampus* **4**, 661–682 (1994).
11. Treves, A. & Rolls, E.T. Computational analysis of the role of the hippocampus in memory. *Hippocampus* **4**, 374–391 (1994).
12. Bakker, A., Kirwan, C.B., Miller, M. & Stark, C.E. Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science* **319**, 1640–1642 (2008).
13. Kirwan, C.B. & Stark, C.E. Overcoming interference: an fMRI investigation of pattern separation in the medial temporal lobe. *Learn. Mem.* **14**, 625–633 (2007).
14. Leutgeb, J.K., Leutgeb, S., Moser, M.B. & Moser, E.I. Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science* **315**, 961–966 (2007).
15. Nakazawa, K. *et al.* Requirement for hippocampal CA3 NMDA receptors in associative memory recall. *Science* **297**, 211–218 (2002).
16. Alvarez, P. & Squire, L.R. Memory consolidation and the medial temporal lobe: a simple network model. *Proc. Natl. Acad. Sci. USA* **91**, 7041–7045 (1994).
17. McClelland, J.L., McNaughton, B.L. & O'Reilly, R.C. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* **102**, 419–457 (1995).
18. Hoffman, K.L. & McNaughton, B.L. Coordinated reactivation of distributed memory traces in primate neocortex. *Science* **297**, 2070–2073 (2002).
19. Ji, D. & Wilson, M.A. Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat. Neurosci.* **10**, 100–107 (2007).
20. Sutherland, G.R. & McNaughton, B. Memory trace reactivation in hippocampal and neocortical neuronal ensembles. *Curr. Opin. Neurobiol.* **10**, 180–186 (2000).
21. Foster, D.J. & Wilson, M.A. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* **440**, 680–683 (2006).
22. Karlsson, M.P. & Frank, L.M. Awake replay of remote experiences in the hippocampus. *Nat. Neurosci.* **12**, 913–918 (2009).
23. Rasch, B., Büchel, C., Gais, S. & Born, J. Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science* **315**, 1426–1429 (2007).
24. Rasch, B. & Born, J. Maintaining memories by reactivation. *Curr. Opin. Neurobiol.* **17**, 698–703 (2007).
25. Knutson, B., Adams, C.M., Fong, G.W. & Hommer, D. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J. Neurosci.* **21**, RC159 (2001).
26. Adcock, R.A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B. & Gabrieli, J.D. Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron* **50**, 507–517 (2006).
27. Wittmann, B.C. *et al.* Reward-related fMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron* **45**, 459–467 (2005).
28. Brewer, J.B., Zhao, Z., Desmond, J.E., Glover, G.H. & Gabrieli, J.D. Making memories: brain activity that predicts how well visual experience will be remembered. *Science* **281**, 1185–1187 (1998).
29. Wagner, A.D. *et al.* Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science* **281**, 1188–1191 (1998).
30. Knutson, B. & Wimmer, G.E. Splitting the difference: how does the brain code reward episodes? *Ann. NY Acad. Sci.* **1104**, 54–69 (2007).
31. McClure, S.M., Laibson, D.I., Loewenstein, G. & Cohen, J.D. Separate neural systems value immediate and delayed monetary rewards. *Science* **306**, 503–507 (2004).
32. Lansink, C.S. *et al.* Preferential reactivation of motivationally relevant information in the ventral striatum. *J. Neurosci.* **28**, 6372–6382 (2008).
33. Pennartz, C.M. *et al.* The ventral striatum in off-line processing: ensemble reactivation during sleep and modulation by hippocampal ripples. *J. Neurosci.* **24**, 6446–6456 (2004).
34. Lansink, C.S., Goltstein, P.M., Lankelma, J.V., McNaughton, B.L. & Pennartz, C.M. Hippocampus leads ventral striatum in replay of place-reward information. *PLoS Biol.* **7**, e1000173 (2009).
35. Marshall, L. & Born, J. The contribution of sleep to hippocampus-dependent memory consolidation. *Trends Cogn. Sci.* **11**, 442–450 (2007).
36. Peigneux, P. *et al.* Are spatial memories strengthened in the human hippocampus during slow wave sleep? *Neuron* **44**, 535–545 (2004).
37. Peigneux, P. *et al.* Offline persistence of memory-related cerebral activity during active wakefulness. *PLoS Biol.* **4**, e100 (2006).
38. Nader, K. Memory traces unbound. *Trends Neurosci.* **26**, 65–72 (2003).
39. Anderson, M.C. & McCulloch, K.C. Integration as a general boundary condition on retrieval-induced forgetting. *J. Exp. Psychol. Learn. Mem. Cogn.* **25**, 608–629 (1999).
40. Tse, D. *et al.* Schemas and memory consolidation. *Science* **316**, 76–82 (2007).
41. Heckers, S., Zalesak, M., Weiss, A.P., Ditman, T. & Titone, D. Hippocampal activation during transitive inference in humans. *Hippocampus* **14**, 153–162 (2004).
42. Shohamy, D. & Wagner, A.D. Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron* **60**, 378–389 (2008).
43. Zalesak, M. & Heckers, S. The role of the hippocampus in transitive inference. *Psychiatry Res.* **172**, 24–30 (2009).
44. Kumaran, D., Summerfield, J.J., Hassabis, D. & Maguire, E.A. Tracking the emergence of conceptual knowledge during human decision making. *Neuron* **63**, 889–901 (2009).
45. O'Reilly, R.C. & Rudy, J.W. Computational principles of learning in the neocortex and hippocampus. *Hippocampus* **10**, 389–397 (2000).
46. Okado, Y. & Stark, C.E. Neural activity during encoding predicts false memories created by misinformation. *Learn. Mem.* **12**, 3–11 (2005).

ONLINE METHODS

Subjects. 20 subjects (18–31 years old, 8 male) participated in the fMRI experiment after informed written consent was obtained in accord with the Stanford University Institutional Review Board. All subjects were right-handed, native English speakers and were paid $20 per h plus bonuses based on task performance. An additional five subjects were excluded from analysis. One subject was excluded for excessive movement, one subject was excluded for not understanding the task instructions and three subjects were excluded on the basis of low performance/compliance. Two of these low-performing subjects were characterized by extremely poor performance specifically for low-reward AB pairs, as revealed at post-test (mean recall for low reward pairs = 3.0%, mean recall for high reward pairs = 42.9%); during debriefing, these subjects reported ignoring or making little to no effort to learn the low-reward pairs, relative to high-reward pairs, despite instructions to attempt to learn all pairs. The third low-performing subject was characterized by overall poor performance (13.1% recall for low-reward AB pairs; 11.9% for high-reward pairs).

Materials. All stimuli used in the memory task were color, clipart style pictures of common objects, with the name of each object presented in text below the image. In total, 464 pictures, along with corresponding names, were used in the experiment; hereinafter referred to as items. 16 items were fillers and the remaining items were divided into 16 sets of 28 items each. For each subject, sets were randomly assigned to the various experimental conditions and randomly combined to construct pairs of items.

Procedure and design. The experiment contained four phases: encoding, immediate test (retrieval), reward localizer and post-test. The encoding and immediate test phases took place during fMRI scanning and were divided into eight alternating rounds (encoding, immediate test, encoding, immediate test, etc.). After the encoding and immediate test rounds, subjects completed the monetary incentive delay (MID) task[25] to independently localize regions sensitive to reward. Finally, subjects completed a post-test outside of the scanner.

During encoding rounds, item pairs were presented for 3.5 s (followed by a 0.5-s fixation cross) and represented either AB or AC pairs. AB pairs consisted of novel cues (left-hand, A) items paired with novel associates (right hand, B) items (for example, 'watch-sink'). AC pairs consisted of repeated cues (A items) paired with novel associates (C items) (for example, 'watch-pipe'). Thus, interference between AB and AC pairs was attributable to the common A term. In addition, all pairs were associated with either high ($2.00) or low ($0.10) reward, indicating potential earnings if the pair was later remembered. Reward values were presented for 1.6 s, followed by a variable duration fixation cross (0.4, 2.4 or 4.4 s) and then the presentation of each pair. Half of all AB pairs were associated with high reward and half with low reward; likewise for AC pairs. Of the high-reward AB pairs, one-third were followed by high-reward AC pairs, one-third were followed by low-reward AC pairs and one-third were not followed by an AC; likewise for low reward AB pairs. Each condition contained 28 pairs or triplets. B and C items paired with a given A item did not begin with the same first letter. AB pairs were distributed across rounds 1–7 and AC pairs across rounds 2–8. Thus, AB and AC pairs were intermixed in rounds 2–7. Trials were separated by variable duration null events (0–12 s).

Each immediate test round mirrored the immediately preceding encoding round; each of the pairs encoded in encoding round *n* were tested in the immediately following test round *n*. Each test trial lasted 3.5 s (followed by 0.5-s fixation cross) and consisted of the presentation of an A item in the same left-hand position as seen during the encoding round, along with a '?' where a B or C item had previously appeared. Subjects attempted to recall the corresponding B or C item aloud or responded 'don't know'. Responses were recorded via microphone. Subjects were instructed that some left-hand items (A) would be associated with multiple right-hand items (B and C) and subjects were to retrieve the most recent associate in such cases (that is, C). Notably, if an A item was associated with both a B and C item, the B item was always presented in encoding round *n* and the C item was always presented in encoding round *n* + 1. This meant that each B item was tested immediately before a corresponding C item would be studied. Earnings were determined by randomly selecting 10% of all immediate test phase trials and awarding subjects either $2.00 or $0.10, as indicated during study, for each pair successfully recalled. Subjects were informed of the basis for reward before

beginning the experiment, but were not reminded of reward values during the immediate test rounds, nor did they receive feedback during these rounds.

Because the immediate test phases were conducted concurrent with fMRI data collection, subjects' verbal responses were recorded via microphone and were later coded for accuracy. Data from one subject were not recorded because of a technical problem with the microphone, resulting in this subject not contributing data to the immediate test phase analysis. For any responses that were ambiguous, multiple experimenters coded the response, blind to the experimental condition associated with each trial. In general, the vast majority of test trials were coded with reasonably high confidence. Notably, because the post-test that yielded the principal behavioral data driving the fMRI analyses was conducted outside of the scanner, data from this critical test were not subject to any coding ambiguities.

The MID task was similar to versions described previously[25]. On each trial, subjects were presented with one of four reward values: +$2.00, +$0.10, –$0.10 or –$2.00 (high positive, low positive, low negative and high negative, 15 trials per condition). On each trial, the reward value was presented for 1.6 s, followed by a variable duration fixation cross (0.4, 2.4 or 4.4 s). After the fixation cross, a triangle was briefly presented (range of 150–700 ms); subjects were instructed to press a key on a button box while the triangle was on the screen. Next, a feedback message ('hit' or 'miss') was presented for 800 ms, followed by another fixation cross for 100–650 ms. If subjects successfully responded while the triangle was on the screen, they received feedback indicating that the trial was a 'hit'; if they failed to respond when the triangle was on the screen, they received feedback indicating that the trial was a 'miss'. For positive reward value trials, subjects earned the indicated money if the trial was a hit; for negative reward trials, hits allowed subjects to avoid losing the indicated money. Misses were associated with either no gain (positive reward trials) or a loss (negative reward trials). Thus, it was always in the subject's best interest to respond while the triangle was on the screen.

To approximately equate MID performance across subjects, we used an adaptive algorithm that dynamically adjusted the duration of the triangle presentation as a function of subject performance. Four independent 'trains' were used, representing the four different reward values. For each train, the target accuracy was 66.0% and the duration of the triangle, which was always initialized to 300 ms, was adjusted trial by trial, depending on whether the subject's running accuracy for that train was above or below target accuracy. For example, if overall accuracy in the high-positive condition after trial *n* was equal to 50%, then the triangle duration for trial *n* + 1 in the high-positive condition was lengthened (making the trial easier). In this manner, the triangle duration was shortened or lengthened by 25-ms increments, depending on whether performance in that condition was above or below target accuracy, with the exceptions that if a subject's running accuracy in a condition was below target accuracy and the last two trials in that condition were hits or if a subject's running accuracy in a condition was above target accuracy and the last two trials were misses, then the duration on the next trial in that condition remained unchanged. This algorithm effectively ensured that net earnings were positive.

Following the MID task, a high-resolution anatomical image was collected and subjects then exited the scanner and completed the last phase of the experiment, the post-test. The post-test was a surprise to subjects, but was similar to the immediate test phase; subjects were presented with A items and asked to recall corresponding associates. However, only B items were tested. If multiple associates (B and C items) had been studied with a given A item, subjects were to recall the first associate. To facilitate recall of the B items, each A item was accompanied by the first letter of the corresponding B item, meaning that B items were uniquely cued. Subjects were informed that their performance in this phase would not affect their earnings (which were based on performance during the immediate test rounds in the scanner).

fMRI methods and procedures. Scanning was conducted at the Stanford University Lucas Center on a 3.0T GE Signa MRI system (GE Medical Systems). Functional images were obtained using a T2*-weighted two-dimensional gradient echo spiral-in/out pulse sequence[47] (repetition time = 2 s, echo time = 30 ms, flip angle = 75°, 30 slices, 3.4 × 3.4 × 4 mm, axial oblique sequential acquisition). We collected 17 functional scans: eight encoding (1,364 volumes), eight immediate test (896 volumes) and one MID (199 volumes). In the encoding and immediate test rounds, inter-trial intervals consisted of 'null' events, during which subjects indicated via button press the left or right direction of rapidly presented arrows.

During the MID task, null events consisted of fixation crosses. All null event durations were pseudo-randomized to optimize design efficiency.

Image preprocessing and data analysis were performed using SPM5 (Wellcome Trust Centre for Neuroimaging, University College London). Functional data were corrected for slice-timing and head motion. Structural images were co-registered to functional images and segmented into gray matter, white matter and cerebrospinal fluid. Gray-matter images were stripped of remaining skull and normalized to a gray-matter MNI template image. Normalized gray-matter images were used for normalization of the structural and functional images. Images were re-sampled to 3-mm cubic voxels and smoothed with a Gaussian kernel (8 mm at full-width half-maximum).

Statistical analyses. Data were analyzed under the assumptions of the general linear model (GLM). Trials were modeled using a canonical hemodynamic response function and its first-order temporal derivative; scan session was treated as a covariate. Study, test and MID data were modeled separately. In addition to the primary GLM (**Supplementary Table 6**), three additional GLMs were constructed to test specific hypotheses (**Supplementary Tables 10**, **12** and **14**). For all GLMs, linear contrasts were used to obtain subject-specific estimates for each effect of interest, which were then entered into a second-level analysis, treating subject as a random effect and using a one-sample t test against a contrast value of zero at each voxel. Unless otherwise noted, a threshold of $P < 0.001$, uncorrected, was used for group-level contrasts. Small volume corrected P values are reported for the main analyses. Contrast maps were overlaid on a mean anatomical image. Unless otherwise noted, ROI analyses were performed by extracting beta values from all significantly active voxels in a 6-mm radius of local maxima.

47. Glover, G.H. & Law, C.S. Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. *Magn. Reson. Med.* **46**, 515–522 (2001).