

# Cognitive neuroscience: Forgetting of things past

Anthony D. Wagner\*<sup>†‡</sup> and Lila Davachi\*

**Recent functional imaging and electrophysiological results indicate that failure to remember experiences can result from a decreased recruitment of encoding processes that build effective memories and an increased recruitment of alternative mechanisms that may impair effective learning.**

Addresses: \*Department of Brain and Cognitive Sciences and <sup>†</sup>Center for Learning and Memory, NE20-463, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>‡</sup>Massachusetts General Hospital, Nuclear Magnetic Resonance Center, USA. E-mail: awagner@psyche.mit.edu

*Current Biology* 2001, 11:R964–R967

0960-9822/01/\$ – see front matter

© 2001 Elsevier Science Ltd. All rights reserved.

*Mental states of every kind — sensations, feelings, ideas — which were at one time present in consciousness and then have disappeared from it, have not with their disappearance absolutely ceased to exist. Although the inwardly-turned look may no longer be able to find them, nevertheless they have not been utterly destroyed and annulled, but in a certain manner they continue to exist, stored up, so to speak in memory. Ebbinghaus (1885)*

Memory provides a bridge between past and present. Through memory, past sensations, feelings and ideas that have dropped from conscious awareness can be subsequently recovered to guide current thought and action. In this manner, memory allows us to locate our car in the parking lot at the end of the day, or guides us to avoid retelling the same joke to the same friend. Continuity through memory critically requires that past experiences ‘continue to exist’ by being effectively ‘stored up’ or encoded in memory [1]. Many of memory’s frailties, including instances of forgetting, emerge from failures at this initial learning stage [2]. For this reason, cognitive neuroscientists have sought to determine the neurocognitive processes that give rise to effective encoding, as well as those that may lead memory astray.

Specifically, what are the storage computations that transform an experience into a durable memory that can support subsequent remembrance? Are there processes which, when recruited during an experience, increase the probability of later forgetting the experience? Answers to these and related questions are beginning to emerge through the application of event-related functional magnetic resonance imaging (fMRI) and single-unit electrophysiological recordings in humans, both of which provide on-line measures of neural activation engaged during event processing. Critically, these

approaches allow researchers to peer into the human brain as it is in the throes of building memories, thus bearing witness to the encoding computations that influence later remembering or forgetting.

One powerful approach to exploring how memories are encoded is to measure the neural response during an experience and ask whether the magnitude of the response is correlated with later memory, either remembering or forgetting (Figure 1) [3,4]. Over the past few years, fMRI and electrophysiological studies have revealed positive correlations between prefrontal, posterior neocortical and medial temporal lobe (MTL) responses and later remembering [5,6]. For example, greater activation in the left ventrolateral prefrontal and posterior parahippocampal cortices during semantic processing of words was associated with an increased probability of remembering on a later memory test [7,8]. By contrast, greater activation in right prefrontal and bilateral MTL regions during the processing of complex visual stimuli was associated with superior subsequent memory for these visuospatial experiences [9,10].

Such positive correlations have informed our models of encoding, pointing to interactions between prefrontal control processes, posterior neocortical representations and MTL binding mechanisms [11–13]. Although attention has been focused on these activation patterns because they suggest that increased recruitment of specific processes results in more effective learning, such computations may be but part of the encoding story. Indeed, new fMRI data published recently in *Current Biology* by Otten and Rugg [14], in conjunction with a recent single-unit electrophysiological study [15] and reanalyses of prior fMRI results [7,16], suggest that to understand encoding, attention also must be afforded to processes that are negatively correlated with remembering — that is, processes that predict subsequent forgetting.

In their study, Otten and Rugg [14] demonstrated that, to the extent that activation magnitude at encoding is related to later memory performance, greater activation need not always predict superior memory; rather, some instances of greater activation are associated with increased subsequent forgetting. This important observation emerged from a further analysis of data from two prior fMRI experiments [17,18], with the analysis targeting neural regions that were more active during the semantic processing of words later forgotten compared to those later remembered. In both experiments, increased forgetting followed greater responses in posterior cingulate, medial parietal, bilateral inferior parietal and bilateral dorsolateral prefrontal cortices.

Table 1

**Foci positively correlated with forgetting in [7].**

Anatomical region	Coordinates			~BA
<b>L medial prefrontal/Middle prefrontal</b>	<b>-21</b>	<b>31</b>	<b>34</b>	<b>8,9</b>
<b>Precuneus/Posterior cingulate</b>	<b>3</b>	<b>-43</b>	<b>40</b>	<b>31,7</b>
<i>Medial parietal/Precuneus*</i>	<i>-6</i>	<i>-65</i>	<i>34</i>	<i>31,7</i>
<i>R superior prefrontal</i>	<i>21</i>	<i>53</i>	<i>21</i>	<i>10</i>
<i>Posterior cingulate</i>	<i>12</i>	<i>-43</i>	<i>40</i>	<i>31,23</i>
<i>Medial prefrontal</i>	<i>6</i>	<i>44</i>	<i>15</i>	<i>9</i>
<i>R inferior parietal*</i>	<i>50</i>	<i>-49</i>	<i>34</i>	<i>40</i>
<i>L posterior insula</i>	<i>-31</i>	<i>-37</i>	<i>18</i>	
<i>Precentral</i>	<i>9</i>	<i>-27</i>	<i>75</i>	<i>4</i>
<i>R inferior parietal*</i>	<i>56</i>	<i>-43</i>	<i>34</i>	<i>40</i>
<i>Precuneus/Posterior cingulate</i>	<i>-6</i>	<i>-40</i>	<i>40</i>	<i>31,7</i>
<i>R inferior parietal*</i>	<i>50</i>	<i>-43</i>	<i>46</i>	<i>40</i>

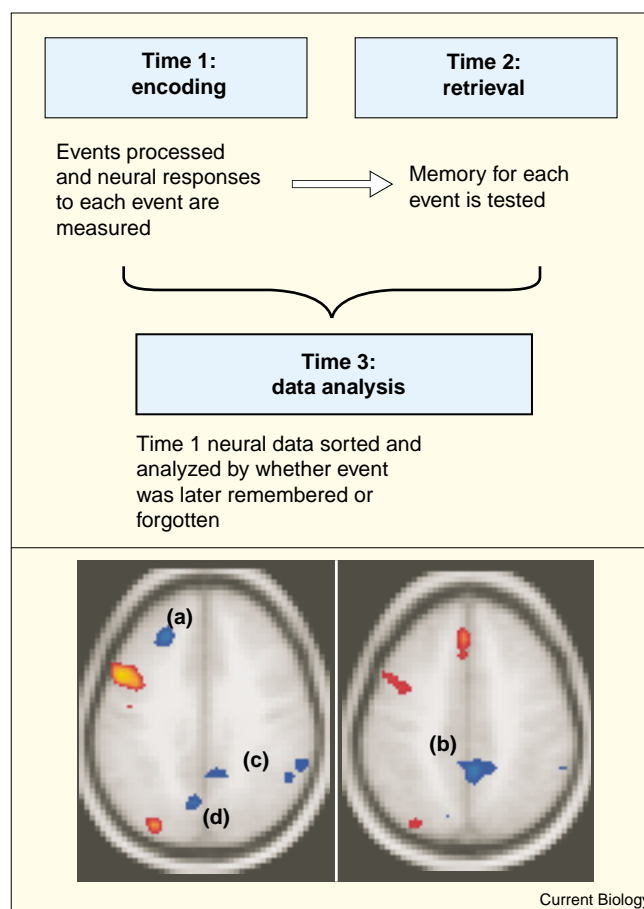
Regions in bold significant at  $P < .001$  (uncorrected); regions in italics significant at  $P < .01$  (uncorrected). ~BA, approximate Brodmann's areas; L, left; R, right; \*similar foci were observed by Otten and Rugg [11].

These subsequent forgetting effects suggest that increased recruitment of some neural computations may actually impair effective memory encoding, perhaps having as significant an impact on later remembrance as do processes that foster effective learning.

Before Otten and Rugg's observations [14], nearly all event-related fMRI reports of subsequent memory effects either failed to observe positive correlates of forgetting [9] or did not explicitly note whether such correlates were observed [10,16–21], with one exception [7]. In this last study, where words were processed in a semantic manner as in Otten and Rugg's experiments, subsequent forgetting effects were observed in the precuneus and in a left prefrontal region that fell medial to Otten and Rugg's dorsolateral response (Table 1). When we reconsidered these data at a more lenient statistical threshold ( $P < .01$  rather than  $P < .001$ ), the exploratory reanalysis revealed predictors of forgetting in additional structures, including right inferior and medial parietal regions that approximated those observed by Otten and Rugg (Figure 1 and Table 1); the medial prefrontal response now extended more laterally towards middle frontal cortex (Figure 1).

Motivated by Otten and Rugg's observations, we also conducted an exploratory reanalysis of data from a recent study of subsequent memory following rote phonological rehearsal of word triplets [16]. Although the orienting task and number of events per trial differed in this study relative to the single-word semantic processing studies noted above,

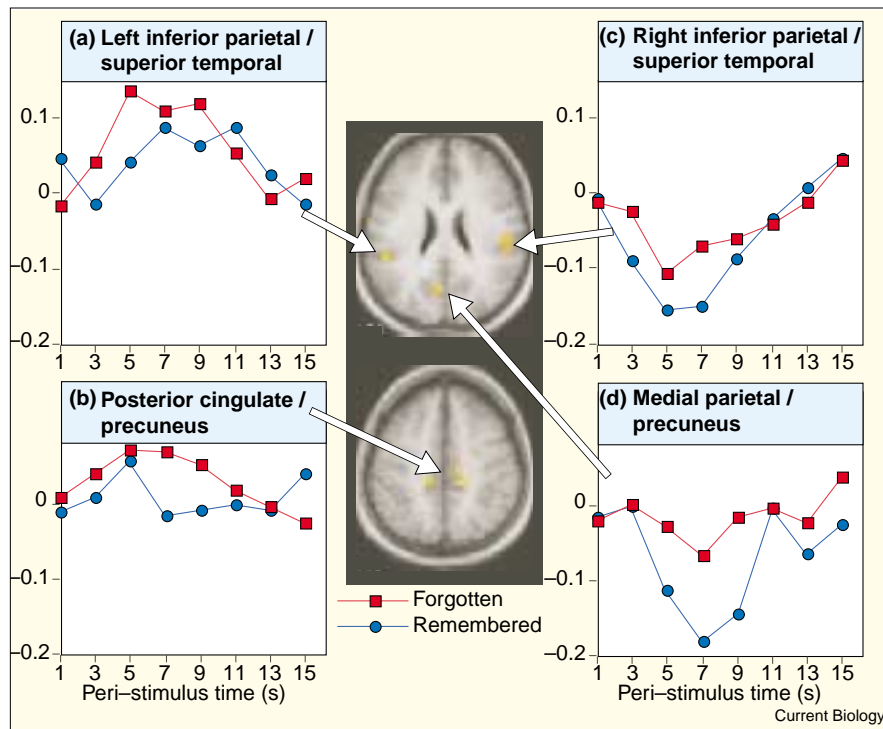
Figure 1



Schematic of the subsequent-memory paradigm and illustrative fMRI results. Top: the paradigm entails recording neural activation during event processing and analyzing these neural data based on behavioral measures of subsequent remembering or forgetting of the events. Bottom: neural regions that showed greater activation during semantic processing of words subsequently forgotten compared to subsequently remembered (blue), and vice versa (red). Activation was positively correlated with forgetting, at a statistical threshold of  $P < 0.01$  (uncorrected), in: (a) left medial prefrontal/middle prefrontal, (b) precuneus/posterior cingulate, (c) right inferior parietal and (d) precuneus/medial parietal cortices (see Table 1 for a complete list of foci; data are from [7]).

nevertheless activation foci that approximated those from Otten and Rugg [14] and Wagner *et al.* [7] were observed. Specifically, positive correlates of forgetting were found in posterior cingulate, bilateral inferior parietal and medial parietal/precuneus cortices; the parietal responses were somewhat ventral to those in the semantic processing studies (Figure 2). Such subsequent forgetting effects likely were overlooked in earlier fMRI studies because of an assumption that processes that influence encoding will be revealed by positive correlations between later memory ability and neural activation during learning [14]. Otten and Rugg's findings, in conjunction with the presently

Figure 2



Regions demonstrating greater activation during phonological processing of items later forgotten compared to those later remembered, and the corresponding event-related hemodynamic responses (in percent signal change). Greater activation was observed in multiple regions, including: (a) left inferior parietal/superior temporal ( $-54, -39, 24$ ), (b) posterior cingulate/precuneus ( $12, -18, 48$ ), (c) right inferior parietal/superior temporal ( $57, -30, 21$ ), and (d) medial parietal/precuneus ( $-6, -66, 27$ ) cortices. The first two regions demonstrated a subsequent forgetting effect superimposed on above baseline activation, whereas this effect in the latter two regions was superimposed on below baseline activation; at present it is unclear whether above and below baseline hemodynamic responses warrant distinct interpretations. Statistical threshold of  $P < 0.01$  (uncorrected); data are from [16].

reported reanalyses of our prior fMRI data, bring into bold relief that this assumption is likely to be incorrect.

Given that subsequent forgetting effects are reliable, the next challenge is to specify the nature of the processes that, when recruited, negatively affect learning. Full consideration of this issue awaits further investigation, as the above studies were not designed to test specific hypotheses about the mechanisms yielding forgetting. Nevertheless, a number of candidate hypotheses appear to warrant brief mention. First, subsequent forgetting may arise because of the diversion of neurocognitive resources away from processes that yield effective encoding. For example, Otten and Rugg [14] proposed that activations positively correlated with forgetting may reflect two possible forms of resource diversion: the devotion of greater resources to the process of selecting task-relevant semantic knowledge during word processing; or the devotion of greater resources to executing a process switch, either between task sets or between stages of processing within a given task. Both accounts suggest that, if greater resources are allocated to a particular goal-appropriate stage of event processing, this could limit resources for other stages that may foster encoding.

The diversion of cognitive resources could also emerge through execution of goal-inappropriate processes, including: attending to task-irrelevant stimulus features that, when

encoded, are not as effective for later item recognition as are semantic codes [22,23]; or diverting attention away from the stimulus and towards other irrelevant thoughts [24]. In the former case, encoding of stimulus-specific traces that do not facilitate item recognition may occur; evidence of such encoding might emerge were the subsequent memory test specifically designed to probe memory for the attended irrelevant features [25].

A second possibility is that subsequent forgetting arises because of encoding processes that yield undifferentiated traces. For example, the hippocampal component of the MTL memory system is posited to mediate encoding through pattern separation processes that give rise to sparse hippocampal representations [26]. Failure to generate differentiated traces may yield interference and forgetting as a result of representational overlap between events.

Although fMRI studies have typically observed positive correlations between MTL activation and later remembering [6], a recent single-unit electrophysiological study of hippocampal subsequent memory effects yielded intriguing results [15]. In particular, of 13 hippocampal neurons that predicted subsequent memory in that study, 10 had a higher firing rate during the encoding of events later forgotten. One might speculate that these higher firing rates reflect ineffective or poor pattern separation during learning, thus resulting in interference and subsequent forgetting.

With regard to subsequent forgetting effects observed in the neocortex by fMRI, it is unclear whether sparse neocortical traces, which may be reflected in reduced fMRI signal [27,28], are effective encoding outcomes.

As Ebbinghaus noted well over a century ago [1], when our current experiences become the past, it is not that they are 'utterly destroyed and annulled'. Rather, through memory's ability to store long-term traces of our experiences, records of the past are built thus permitting the past to reemerge, through retrieval, at some future time. The presently considered fMRI and electrophysiological data reflect advances in understanding how effective memories are built, as well as highlight the uncertainties that remain for subsequent investigation, including further delineation of the processes that increase forgetting of things past.

#### Acknowledgements

Supported by the NIDCD (DC04466), McKnight Endowment Fund for Neuroscience, Surdna Foundation, and P.E. Newton.

#### References

- Ebbinghaus H: *Memory: A Contribution to Experimental Psychology*. New York: Dover; 1885/1964.
- Schacter DL: *The Seven Sins of Memory: How the Mind Forgets and Remembers*. Boston and New York: Houghton-Mifflin; 2001.
- Sanquist TF, Rohrbaugh JW, Syndulko K, Lindsley DB: **Electrophysiological signs of levels of processing: Perceptual analysis and recognition memory.** *Psychophysiology* 1980, **17**:568-576.
- Paller KA, Kutas M, Mayes AR: **Neural correlates of encoding in an incidental learning paradigm.** *Electroencephalogr Clin Neurophysiol* 1987, **67**:360-371.
- Wagner AD, Koutstaal W, Schacter DL: **When encoding yields remembering: Insights from event-related neuroimaging.** *Phil Trans R Soc London Ser B* 1999, **354**:1307-1324.
- Paller KA, Wagner AD: **Observing the transformation of experience into memory.** *Trends Cognit Sci*, submitted.
- Wagner AD, Schacter DL, Rotte M, Koutstaal W, Maril A, Dale AM, Rosen BR, Buckner RL: **Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity.** *Science* 1998, **281**:1188-1191.
- Henson RNA, Rugg MD, Shallice T, Josephs O, Dolan RJ: **Recollection and familiarity in recognition memory: An event-related functional magnetic resonance imaging study.** *J Neurosci* 1999, **19**:3962-3972.
- Brewer JB, Zhao Z, Desmond JE, Glover GH, Gabrieli JD: **Making memories: brain activity that predicts how well visual experience will be remembered.** *Science* 1998, **281**:1185-1187.
- Kirchhoff BA, Wagner AD, Maril A, Stern CE: **Prefrontal-temporal circuitry for novelty encoding and subsequent memory.** *J Neurosci* 2000, **20**:6173-6180.
- Moscovitch M: **Memory and working-with-memory: a component process model based on modules and central systems.** *J Cognit Neurosci* 1992, **4**:257-267.
- Buckner RL, Kelley WM, Petersen SE: **Frontal cortex contributes to human memory formation.** *Nat Neurosci* 1999, **2**:311-314.
- Wagner AD, Maril A, Schacter DL: **Interactions between forms of memory: When priming hinders new learning.** *J Cognit Neurosci* 2000, **12**:S2:52-60.
- Otten LJ, Rugg MD: **When more means less: neural activity related to unsuccessful memory encoding.** *Curr Biol* 11:1528-1530.
- Cameron KA, Yashar S, Wilson CL, Fried I: **Human hippocampal neurons predict how well word pairs will be remembered.** *Neuron* 2001, **30**:289-298.
- Davachi L, Maril A, Wagner AD: **When keeping in mind supports later bringing to mind: neural markers of phonological rehearsal predict subsequent remembering.** *J Cognit Neurosci*, in press.
- Otten LJ, Henson RN, Rugg MD: **Depth of processing effects on neural correlates of memory encoding: Relationship between findings from across- and within-task comparisons.** *Brain* 2001, **124**:399-412.
- Otten LJ, Rugg MD: **Task-dependency of the neural correlates of episodic encoding as measured by fMRI.** *Cereb Cortex*, in press.
- Canli T, Zhao Z, Brewer J, Gabrieli JD, Cahill L: **Event-related activation in the human amygdala associates with later memory for individual emotional experience.** *J Neurosci* 2000, **20**:RC99.
- Buckner RL, Wheeler ME, Sheridan MA: **Encoding processes during retrieval tasks.** *J Cognit Neurosci* 2001, **13**:406-415.
- Baker JT, Sanders AL, Maccotta L, Buckner RL: **Neural correlates of verbal memory encoding during semantic and structural processing tasks.** *NeuroReport* 2001, **12**:1251-1256.
- Craik FIM, Lockhart RS: **Levels of processing: a framework for memory research.** *J Verbal Learn Verbal Behav* 1972, **11**:671-684.
- Craik FI, Tulving E: **Depth of processing and the retention of words in episodic memory.** *J Exp Psychol Gen* 1975, **104**:268-294.
- Craik FI, Govoni R, Naveh-Benjamin M, Anderson ND: **The effects of divided attention on encoding and retrieval processes in human memory.** *J Exp Psychol Gen* 1996, **125**:159-180.
- Morris CD, Bransford JD, Franks JJ: **Levels of processing versus transfer appropriate processing.** *J Verbal Learn Verbal Behav* 1977, **16**:519-533.
- McClelland JL, McNaughton BL, O'Reilly RC: **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* 1995, **102**:419-457.
- Wiggs CL, Martin A: **Properties and mechanisms of perceptual priming.** *Curr Opin Neurobiol* 1998, **8**:227-233.
- Wagner AD, Koutstaal W, Maril A, Schacter DL, Buckner RL: **Task-specific repetition priming in left inferior prefrontal cortex.** *Cereb Cortex* 2000, **10**:1176-1184.