Predictive, Interactive Multiple Memory Systems

Richard N. Henson* and Pierre Gagnepain

ABSTRACT: Most lesion studies in animals, and neuropsychological and functional neuroimaging studies in humans, have focused on finding dissociations between the functions of different brain regions, for example in relation to different types of memory. While some of these dissociations can be questioned, particularly in the case of neuroimaging data, we start by assuming a ''modal model'' in which at least three different memory systems are distinguished: an episodic system (which stores associations between items and spatial/temporal contexts, and which is supported primarily by the hippocampus); a semantic system (which extracts combinations of perceptual features that define items, and which is supported primarily by anterior temporal cortex); and modality-specific perceptual systems (which represent the sensory features extracted from a stimulus, and which are supported by higher sensory cortices). In most situations however, behavior is determined by interactions between these systems. These interactions reflect the flow of information in both ''forward'' and ''backward'' directions between memory systems, where backward connections transmit predictions about the current item/features based on the current context/item. Importantly, it is the resulting ''prediction error''—the difference between these predictions and the forward transmission of sensory evidence—that drives memory encoding and retrieval. We describe how this ''predictive interactive multiple memory systems'' (PIMMS) framework can be applied to human neuroimaging data acquired during encoding or retrieval phases of the recognition memory paradigm. Our novel emphasis is thus on associations rather than dissociations between activity measured in key brain regions; in particular, we propose that measuring the functional coupling between brain regions will help understand how these memory systems interact to guide behavior. © 2010 Wiley-Liss, Inc.

KEY WORDS: recollection; familiarity; priming; memory systems

INTRODUCTION

The debate over whether recollection and familiarity are qualitatively different memory processes reflects a broader debate over the fractionation of memory, such as, whether there exist distinct memory systems (e.g., Schacter and Tulving, 1994). While multiple memory processes/ systems may be helpful at one level of description—for example, to explain behavioral dissociations between different types of memory test—they may not necessarily have spatially-segregated or independently-operating implementation in the brain. Nonetheless, dissociations

in the pattern of brain activity during various memory tasks, as measured for example by functional magnetic resonance imaging (fMRI), are often cited as further evidence in support of multiple memory processes. These fMRI claims most often relate to the medial temporal lobes (MTL), particularly dissociations between the hippocampus and surrounding perirhinal cortex and parahippocampal gyri (e.g., Diana et al., 2007). Here we evaluate these dissociations critically, at least within the domain of fMRI studies of human recognition memory (see Brown et al., 2010; Montaldi and Mayes, this issue, for evidence from animal and human lesion studies), before proposing a new framework that emphasizes instead associations rather than dissociations between the activity of different brain regions.

We begin by briefly reviewing cognitive and neural theories of recognition memory, before illustrating some criteria for establishing neuroimaging support for multiple memory systems, based on patterns of fMRI response in hippocampus and perirhinal cortex associated with recollection and familiarity. Even if no study has yet met these criteria fully, we nonetheless assume a modal, multiple-memory-systems model in which recollection and familiarity depend on the "forward" and "backward" coupling between systems. We call this model PIMMS, which stands for "predictive interactive multiple-memory systems," where the novel "predictive" and "interactive" aspects refer to the idea that "higher" systems are constantly trying to predict activity in "lower" systems on the basis of past experience. We sketch how PIMMS might explain the common patterns of recollection- vs. familiarity-related activity in hippocampus and perirhinal cortex, before generalizing to more everyday situations, and making predictions for future experiments. While it becomes more difficult to test such fully-interactive models based on behavioral data or even local patterns of fMRI responses, we suggest that they can also be tested by examining changes in the functional coupling between brain regions.

Theories of Recognition Memory

People's ability to recognize a stimulus that they encountered previously has been hypothesized to entail two qualitatively different types of memory: recollection and familiarity (e.g., Mandler, 1980). In keeping with other articles in this special issue, we use these terms to refer to psychological processes that both give rise to conscious memory for past exposure to a stimulus, differing in whether (recollection) or

not (familiarity) that memory includes details of the episodic context surrounding that exposure. Experimental assay of recollection and familiarity includes objective tests of aspects of that prior context (or "source memory"; Johnson et al., 1993) or subjective judgments of context retrieval (such as "remember" vs. "know" judgments; Tulving, 1985). Rather than discuss the pros and cons of such experimental methods (see instead Montaldi and Mayes, this issue; Wixted et al., 2010), we assume here that recollection has occurred when there is evidence of retrieval of at least one type of extrinsic context (e.g., spatial or temporal information, or internally-generated thoughts during prior exposure). Note that we do not include memory for context intrinsic to the stimulus, such as perceptual details not repeated at test (e.g., Staresina and Davachi, 2008; cf. Wais et al., 2008), which we hypothesize are bound within a semantic memory system (see later).

When fitting behavioral data from recognition memory paradigms, assumptions are needed about how recollection and familiarity cooccur or even interact. In one popular model (see Yonelinas et al., 2010), recollection and familiarity are assumed to be independent. In such "dual-process" models of recognition memory, familiarity is normally assumed to be a continuous signal, on which people place response criteria to make their recognition decision, consistent with signal detection theory (SDT, Green snd Swets, 1966). Recollection on the other hand can be modeled as a discrete probabilistic event (a "high-threshold" model) or as a second (orthogonal) continuous signal (in two-dimensional SDT theories, e.g., Rotello et al., 2004). Others however have argued that the behavioral data are sufficiently modeled by one-dimensional SDT (see e.g., Wixted et al., 2010). Some of these researchers take this sufficiency to indicate that there is only one type of memory process underlying recognition memory ("single-process" theories, such as some "global matching" models, Humphreys et al., 1989), such that recollection and familiarity are regarded as quantitatively rather than qualitatively different. It is important to note however that one-dimensional SDT models do not necessarily reject the idea of qualitatively different memory processes; they assume simply that, at least for a basic old/new decision, all information retrieved from memory is combined into a single "strength of evidence" in order for people to make a decision. Nonetheless, one take on this ongoing debate between single- and dual-process models is that behavioral data from the recognition memory paradigm are insufficient to resolve this debate (particularly if SDT is generalized to allow unequal signal and noise variances or even non-Gaussian distributions; see, e.g., Shimamura, 2010), and one should instead seek convergent evidence from other memory paradigms and other types of data, such as those from neuroimaging (Henson, 2006).

Here we distinguish between memory processes and memory systems, where we take the latter to entail spatially-distinct neural implementations in the brain. In this sense, dual process models of recognition memory do not imply multiple memory systems. Recollection and familiarity, for example, could arise from different computations within the same brain region (e.g., Greve et al., 2010), or, as later, from different types of interac-

tion between the same set of brain regions. Conversely though, spatially-distinct neural systems (for the same type of memoranda) would seem to imply different memory processes. Hence, evidence from neuroimaging (or other brain data) that different brain regions are differentially involved in experimental conditions presumed to differ only in terms of recollection and familiarity would support dual-process models of recognition memory.

Several researchers have mapped multiple memory systems onto the brain, focusing on the MTL. Some are based on specific computational principles (e.g., Norman and O'Reilly, 2003); others are verbal accounts (e.g., Aggleton and Brown, 1999). A common theme is that recollection arises from a contextual/relational/pattern-separating system in the hippocampus, whereas familiarity arises from item-based/pattern-completing cortical systems, such as the perirhinal cortex (see, e.g., Norman, 2010; though see also Squire et al., 2007). Before proposing our own modified multiple memory system view, we pause to consider the human neuroimaging evidence for multiple memory systems.

Neuroimaging Evidence for Multiple Memory Systems

To infer from neuroimaging data that more than one process is engaged in a memory task, Henson (2005) argued that one must show a "qualitative" difference in the activity of two more brain regions across three or more experimental conditions. The general idea of a qualitative difference is illustrated in Figure 1. Assume that one measures the mean event-related fMRI response to three types of trial: (1) correct recognition of a studied item together with correct retrieval of its study context (C+); (2) correct recognition of a studied item but failure to retrieve its study context (C−); and (3) correct rejections (CR) of unstudied items. Assume further that one obtains these data from two, independently-defined brain regions: perirhinal cortex and hippocampus.

The null hypothesis here is that the three conditions differ only along a single dimension of "memory strength," with a relative ordering of activity CR < C− < C+ (i.e., a single-process model). An alternative hypothesis is that the conditions differ in terms of recollection (which is greater in C+ than C−) and familiarity (which is greater for C− than CR). However, the precise mapping between these psychological processes (memory strength/recollection/familiarity) and neural activity in each brain region (let alone the mapping between neural activity and the fMRI signal) is unknown; it is unlikely to be linear, for example. Therefore, in order to make any type of inference from fMRI data, we make the minimal assumption that this mapping is monotonic (Henson, 2005).

The pattern of data at the top of Figure 1A is not sufficient to reject the null hypothesis, i.e., insufficient to infer separate processes of recollection (in the hippocampus) and familiarity (in the perirhinal cortex). This is simply because the two brain regions may have different mappings between memory process and fMRI signal; though both monotonic, the mapping may "saturate" at lower levels of memory strength in the perirhinal
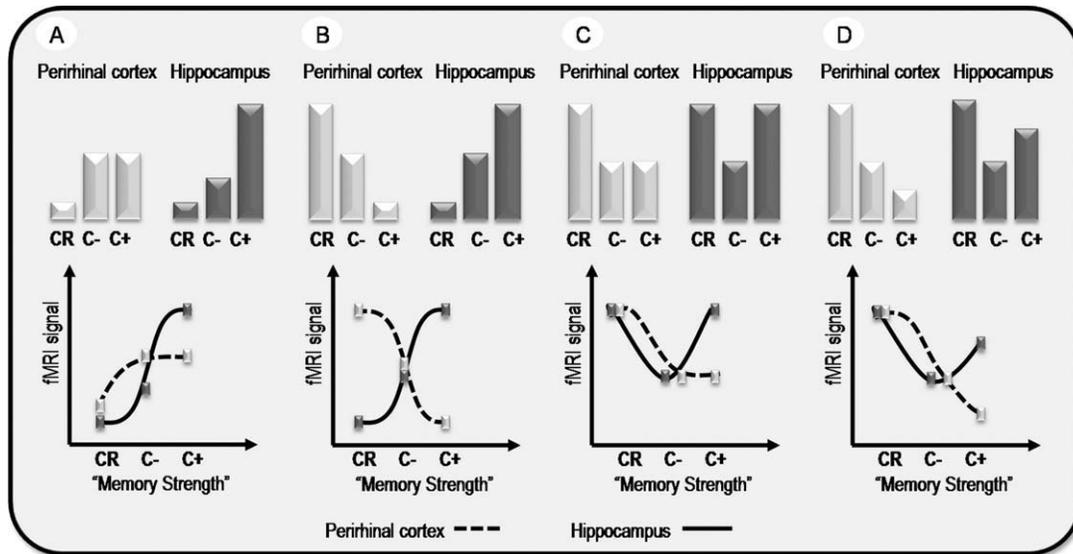
FIGURE 1. Hypothetical patterns of fMRI signal (top row) in perirhinal cortex and hippocampus across three types of trial during the test phase of a recognition memory experiment: correct rejections (CR) of unstudied items; correct recognition of studied items without evidence of any contextual retrieval (C−), analogous to familiarity; and correct recognition of studied items with evidence of contextual retrieval (C+), analogous to recollection. Assuming a monotonic relationship between fMRI signal and degree of memory strength (bottom row), only the patterns in Panels C and D are sufficient to challenge single-system/process models of recognition memory (see text for explanation).

cortex than in the hippocampus (bottom of Fig. 1A; see Henson, 2006, for further elaboration, and Squire et al., 2007, for a similar argument). It is also possible that increases in memory strength can produce decreases in the fMRI signal; for example, if the neural basis of familiarity were reduced responsiveness (e.g., Brown and Xiang, 1998; Brown et al., 2010). In this case, even a cross-over interaction between the three conditions and two regions (Fig. 1B) would not reject the null hypothesis. Indeed, the hippocampus and perirhinal cortex may be functionally-coupled in a reciprocal fashion, whereby increased neural activity in one inhibits activity in the other.

But what about the pattern in Figure 1C? This pattern cannot be explained by a monotonic relationship between memory strength and fMRI signal. Indeed, one common explanation for this pattern is that the hippocampus is involved both in retrieving context, explaining why C+ > C−, and in encoding new items/contexts (even during a recognition test), explaining why CR > C− (e.g., Buckner et al., 2001; Stark and Okado, 2003). Another possibility is that this pattern reflects a spatial averaging over distinct subfields within the hippocampus that are differentially involved in encoding and retrieval, but beyond than the typical resolution of fMRI (Carr et al., 2010). Later, we propose an alternative explanation in terms of contextual-competition in our PIMMS framework.

One might wonder why the data from the perirhinal cortex are relevant to the above argument; with the nonmonotonic pattern in the hippocampus being sufficient to reject the null hypothesis of a single dimension of memory strength. This is only true if the null hypothesis specifies the relative ordering of the C+, C−, and CR conditions; otherwise, one could still assume a single dimension, but where the conditions were

ordered such that C− < C+ < CR, whereby the pattern in Figure 1C reduces back to an inconclusive pattern like that in Figure 1A. In the more general case, a pattern like that in Figure 1D is necessary to rule out any single-process model. This case corresponds to a "reversed association" (Dunn and Kirsner, 1988), where there is a negative-relationship between two conditions (C− and C+), concurrent with a positive-relationship across two other conditions (C−/C+ vs. CR). This pattern requires the data from at least two brain regions, because then there is no way to reorder the three conditions consistently across the two regions and maintain monotonic psychological-fMRI mappings within each region.

In fact, we are not aware of any published fMRI study that has reported a pattern like that in Figure 1C (or 1D), when the regions have been defined independently (i.e., by anatomical criteria, separate functional data, or orthogonal contrasts on the same functional data). The studies that have reported similar patterns at retrieval, including our own, have tended to perform post hoc tests across regions that were defined after searching for significant correlated effects, which raises the question of statistical bias (Henson, 2006).

A New Multiple Memory Systems Model: From SPI to PIMMS

Even if we are not aware of any neuroimaging study to date that has found a qualitative difference in MTL activity associated with recollection and familiarity, according to the above criteria, we propose on the basis of other considerations (e.g., neuropsychological, evolutionary, comparative, and computational, Schacter and Tulving, 1994) that there are multiple

A) SPI (and MIM*)

B) PIMMS
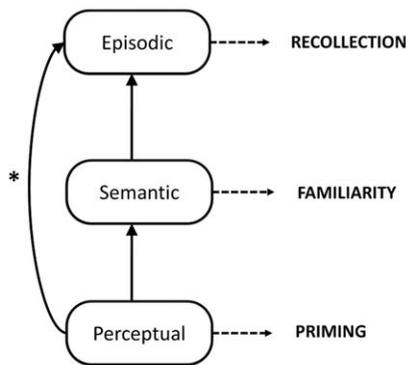
- - - → **Feed-back predictions**
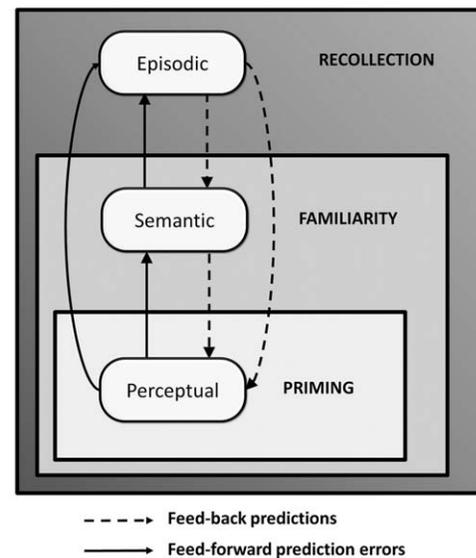—→ **Feed-forward prediction errors**

FIGURE 2.   Tulving's "serial-parallel-independent" (SPI) model of encoding, storage and retrieval of memories (and its extension in the "multiple input model," MIM), and its relationship to recollection, familiarity and priming (Panel A), in contrast to the new predictive interactive multiple memory system framework (PIMMS) proposed here (Panel B), where encoding, storage and retrieval are better described as "interactive-parallel-interactive," and where recollection and familiarity entail interactions between multiple memory systems.

memory systems in the human brain. However, given the high-level of anatomical connectivity between brain regions, we propose that there is a high-degree of interaction between memory systems in most situations, which can make their dissociation and identification difficult.

Perhaps the "modal" multiple memory system model is the serial-parallel-independent (SPI) model developed by Tulving and Gazzaniga (1995)—Figure 2A. This model assumes at least three memory systems: an episodic system, a semantic system, and one or more perceptual systems. The name SPI refers to the relationship between these systems during encoding, storage and retrieval: viz., that encoding is a serial process in which a stimulus is processed first by a perceptual system, then the semantic system and finally the episodic system; storage occurs in parallel, in that this processing may leave separate memory traces in each system; and retrieval is independent, in that such memory traces can be accessed independently (depending on the retrieval test, retrieval cues, etc). According to SPI, recollection is the process of retrieval from the episodic system, while familiarity is the process of retrieval from the semantic system (and the process of retrieval from a perceptual system is priming).

A modification of the SPI model was later proposed by Graham et al. (2000). They proposed that there is a direct pathway between perceptual systems and the episodic system that "bypasses" the semantic system. They argued that this pathway is necessary to explain why recognition memory can be spared in individuals with Semantic Dementia, despite their compromised semantic system, particularly for perceptually-rich but less meaningful stimuli. This pathway may also explain how recollection could be intact despite impaired familiarity, e.g., following a perirhinal lesion (Bowles et al.,

2007). This "multiple inputs model" (MIM) has since been extended to include other systems (working and procedural memory) and to provide a more comprehensive account of human memory (e.g., of autobiographical memory); see Eustache and Desgranges (2008).

Given our focus on recollection and familiarity in tests of recognition memory for unrelated, single items, we will focus on the three-system model in Figure 2B. We distinguish the systems primarily on their representational content and computational principles.[i] We assume that the main purpose of the

---

[i]Since we later propose that these systems operate according to similar principles, differing only in their representational content (and possibly the type of representation, e.g., sparse or distributed), one could question whether these are really "memory" systems. We accept that they are not different systems in the sense of operating according to separate memory principles; or even memory systems in the sense of being specialized for memory (since we believe plasticity is a general property of brain regions, whose specialization depends instead on the nature of the information that they code; see also Ranganath, 2010; Cowell et al., 2010). Instead, we use the phrase "multiple memory systems" here simply to emphasize that different types of memory (episodic/semantic or recollection/familiarity) arise at different stages of interaction within a hierarchy of brain regions (a better name for PIMMS might be "predictive interactive multiple memory signals"). Note also that we do not further characterize these memory systems in terms of different states of consciousness, as does Tulving (2002). We suggest instead that the nature of conscious states—e.g., explicit vs. implicit memory (Graf and Schacter, 1985)—is an orthogonal issue (see also Slotnick, 2010), which most likely depends on interactions between these memory systems and other neural systems, e.g., prefrontal-parietal networks, Dehaene et al. (2006).

episodic system is to bind items to their episodic context. By "item," we refer to visual objects or auditory tokens that are the focus of attention (entailed, for example, by the nature of a memory task). More precisely, an "item" is a hypothetical cause of the current "stimulus" that impinges on the senses. By "context," we refer loosely to other background information like the spatial environment, one's internal thoughts and emotions, and possibly some form of temporal information. We assume the episodic system is also necessary to encode single events defined by the cooccurrence of two or more unrelated items, for example in associative priming (e.g., Degonda et al., 2005). The primary purpose of the semantic system, on the other hand, is to store combinations of perceptual features that repeatedly cooccur (in certain relationships) in the environment, and which thus define items (e.g., Murray et al., 2007; Cowell et al., 2010). This includes binding together features that are represented within different perceptual systems (sensory modalities), and also "chunking" recurring sets of items into larger item representations. The role of the perceptual systems is to abstract and represent those recurring features from the environment that define items.

To make contact with the neuroimaging data, our second assumption is the hippocampus is a key component of the episodic system, the perirhinal cortex is a key component of the semantic system (though extending to anterior temporal lobes more generally, Patterson et al., 2007) and the more posterior cortices that are specialized for a particular sensory modality comprise the perceptual systems (e.g., the ventral visual pathway in occipitotemporal cortex or the auditory pathway in lateral temporal cortex). Note that this anatomical mapping is tentative (see Fig. 3), and each system may of course include several other interconnected regions (e.g., Aggleton and Brown, 1999), not to mention many brain regions important for other factors that affect memory, like motivation, attention, emotionality, and organization, but which are not considered here. Our conception thus shares many similarities with other recent functional-neuroanatomical models of memory (e.g., Davachi, 2006; Diana et al., 2007; Eichenbaum et al., 2007; Ranganath, 2010). However, our novel addition is that there is not only feedforward of information during encoding (as in the SPI model), but also feedback of information, during both encoding and retrieval. The important consequence is that encoding and retrieval are no longer separate processes within each memory system, but arise from recurrent interactions between all three systems (see later).

Thus we have implicitly defined a hierarchy, with the hippocampal episodic system at its apex (e.g., Mishkin et al., 1998). The role of feedback from one system is to predict the activity in "lower" systems in this hierarchy. For example, a representation of the current context (e.g., of the spatial environment, perhaps represented in parahippocampal regions) is used by the hippocampus to predict items that are likely to appear in that context, i.e., to predict activity in the perirhinal cortex. The role of the feedforward flow of information, on the other hand, is to transmit the difference between such "top-down" predic-
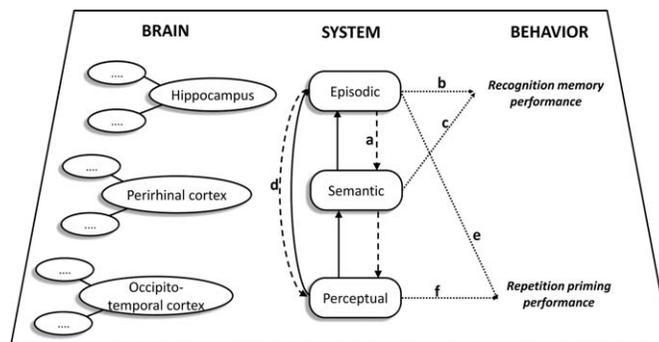

FIGURE 3. The relationship between brain, hypothetical memory system, and behavior in PIMMS. Routes a–f reflect different possible causes of behavioral outcomes in recognition memory or perceptual priming paradigms (see text).

tions and the current "bottom-up" input. The purpose of the recurrent interactions throughout the systems is then to minimize this so-called "prediction error" (Mumford, 1992; Dayan and Hinton, 1996; Rao and Ballard, 1999). For example, the activity of possible representations of the currently-attended visual object in perirhinal cortex is adjusted on the basis of the prediction error passed forward from activity reflecting possible features in occipitotemporal cortex. Perception of a specific object (after a few hundreds of milliseconds of such recurrent processing) occurs when the network of systems settles into a stable state in which the overall prediction error has been minimized.

While the recurrent flow of activity that minimizes prediction error reflects the process of perception, memory arises from subsequent changes in synaptic connections between systems. The nature of these synaptic changes is again such that they reduce prediction error that arises when an item/context recurs (by maximizing the "free energy" of the network as a whole, Friston, 2010). Thus larger residual prediction errors generally entail greater memory storage. Note that the synaptic change necessary for long-term storage is assumed to occur "offline," i.e., possibly minutes to days after encoding a stimulus.

This extension of the neurophysiological/computational idea of "predictive coding" to higher-level memory systems, particularly contextual predictions from episodic to semantic memory systems, is why we call this framework "predictive interactive multiple memory systems" (PIMMS). Furthermore, by basing this framework on the inherent tendency of the brain to attempt to predict its surrounding environment, PIMMS represents a theory of human memory that relates the psychological and neural causes of encoding and retrieval processes to a single principle—the minimization of prediction-error—while maintaining distinctions between different types of representation that distinguish traditional multiple memory systems.

PIMMS is not the first attempt to characterize neural/cognitive processes within a predictive framework (e.g., Bar, 2007;

Buckner, 2010; though see also Bubic et al., 2010). However, we do not use "prediction" here in a general sense, e.g., to refer to anticipation/simulation of future events (cf. Buckner, 2010). Though such temporal predictions may be an important function of episodic memory and the MTL, and a worthy goal for extensions of PIMMS, our use here is restricted to predictions about the nature of the current stimulus (perception) on the basis of past experience (memory): more precisely, predictions about one type of information (e.g., of item representations in perirhinal cortex) from a different type of information in a higher layer (e.g., of context representations in hippocampus). PIMMS is therefore the first attempt to (1) describe how predictions can be framed within an interactive, multiple memory system view, and (2) relate these to behavioral and neuroimaging data on recognition memory.

Behavioral Outcomes According to PIMMS

The neural determinants of behavioral phenomena depend on the specific task demands (presumably implemented by decision mechanisms in frontal/parietal cortices that "read-out" the neural activity from the relevant brain region). Thus in tasks emphasizing visual perceptual decisions, priming will be driven by "read-out" of memory-related changes in activity in occipitotemporal cortex, even if the level of such activity is still affected the feedback of activity from perirhinal and/or hippocampal regions. In tasks emphasizing contextual retrieval, on the other hand, recollection should be driven by read-out of multiple regions, including hippocampus, perirhinal cortex and possibly occipitotemporal cortex.

However, the interactive nature of PIMMS makes it more difficult to isolate different memory systems on the basis of behavioral data alone. This is because it is difficult to distinguish "contamination" of a memory test that is supposed to depend on only one memory system (a "process-pure" test, Jacoby, 1992) from true neural interactions between two or more memory systems. Thus the recognition memory task may be an impure test of episodic memory either because it can be solved by independent contributions from recollection and familiarity (as assumed by the SPI and dual-process models described earlier)—routes b and c in Figure 3—or because "retrieval" of contextual information from the episodic memory system depends on interactions with the semantic memory system (or perceptual systems)—routes a and b in Figure 3. Likewise, a test of repetition priming might include a contribution from a perceptual system, in addition to independent "contamination" by retrieval from the episodic system—routes e and f in Figure 3—or true interactions between perceptual and episodic systems—routes d and f in Figure 3. This high level of interaction may also explain why most dissociations in behavioral performance across different memory tasks are single dissociations, which are easily explained by scaling effects and/or task-specific noise processes (e.g., Berry et al., 2008), (and why the more compelling reversed associations, as defined

earlier, are actually rare, though see Richardson-Klavehn et al., 1999).[ii]

fMRI Correlates of Encoding and Retrieval Related to Familiarity and Recollection According to PIMMS

The "on-line" stabilization process that occurs due to recurrent minimization of prediction error over the timescale of milliseconds (e.g., after presentation of a recognition test cue) is equivalent to the process of "retrieval"; the "off-line" minimization that occurs over the longer timescale due to synaptic change is equivalent to learning/storage. This "storage" stage is not easily detectable with the typical "online" measures of functional neuroimaging. However, the precursors of storage—the process of memory encoding—can be examined by comparing the online activity during presentation of a stimulus as a function of whether that stimulus is later remembered, so-called "subsequent memory" effects (Paller and Wagner, 2002). According to PIMMS, the increased fMRI signal often observed for such subsequent memory effects is expected because larger residual prediction errors (after perception/retrieval has occurred, i.e., when the network has stabilized) entail greater synaptic change, which will tend to lead to more successful encoding. Thus according to PIMMS (and unlike SPI), encoding and retrieval are intimately related in the sense that they both depend on the stabilization of the state of all memory systems into one in which global prediction error is minimized on the basis of stored information.

In general, prediction errors will be greatest when an unexpected stimulus is encountered; such as a novel item (leading to large prediction error in perceptual systems, e.g., occipitotemporal cortex, based on an inability of the semantic system to predict the activity there) or a familiar item in a novel context (leading to large prediction error in perirhinal cortex, based on an inability of the episodic system to predict the activity there). Such novelty is accepted as an important determinant of memory formation, in that it makes evolutionary/metabolic sense not to repeatedly encode information that is already well-predicted by the brain. Below, we focus on the more specific fMRI predictions of PIMMS for encoding and retrieval related to familiarity and recollection in the recognition memory paradigm.

[ii]One empirical observation often taken as evidence of independent retrieval from multiple memory systems (as in SPI) is "stochastic independence": the failure to find significant conditional probability between performance in two memory tests on the same stimulus. However, testing for such stochastic independence requires some care (Poldrack, 1996), and the addition of nonmnemonic, but task-specific noise processes can dramatically reduce the degree of dependence predicted by single-system models (Berry et al., 2008). Moreover, stochastic independence has been found even between two cues within the same memory test (Hayman and Tulving, 1989), questioning this measure as an indicator of independent memory systems.

Given its slow dynamics, the fMRI signal reflects the integration of neural activity over time, and so PIMMS predicts higher fMRI signals when prediction error is greatest, and the network tends to take longer to settle into a stable state.[iii] The reduction in the integrated prediction error over the few hundred milliseconds after stimulus onset has already been proposed as the neural mechanism of "repetition suppression" (Henson, 2003). Repetition suppression refers to the reduced fMRI response for repeated vs. initial presentations of a stimulus, and is sometimes accompanied by an earlier response latency (Gagnepain et al., 2008a), as predicted by a shorter duration of underlying neural activity (Henson et al., 2002). According to PIMMS, repetition suppression in, for example, occipitotemporal cortex can be interpreted in terms of the current visual input being more rapidly "explained away" because of improved predictions for the perceptual features associated with (recently seen) items; improvements caused by synaptic changes triggered by the prediction error that occurred after the initial presentation of the stimulus.

Repetition suppression in occipitotemporal brain regions has been associated with perceptual priming. However, repetition suppression is also found in more anterior temporal regions like perirhinal cortex (Henson et al., 2003), which has been traditionally associated with familiarity instead.[iv] Assuming an absence of strong predictions about a studied or unstudied item based on the current context (e.g., via feedback from hippocampus)—and absence of a strong prediction about the context from the studied item (see below)—then the only difference between studied and unstudied items at test will be the smaller prediction error fed forward from perceptual systems because of synaptic change after study. Therefore a smaller fMRI response is predicted in perirhinal cortex for studied than unstudied items. Note that this greater input to perirhinal cortex for unstudied items, in conjunction with little difference in contextual predictions fed back from hippocampus for studied or unstudied items, in turn entails a greater prediction error fed forward to hippocampus, explaining why CRs can increase fMRI signal relative to C− trials in the hippocampus (Fig. 1C). Indeed, the recurrent activity between perirhinal cortex and hippocampus may entail greater functional coupling between these regions (Fernandez and Tendolkar, 2006), in turn leading to encoding of CRs with the test context (Buckner et al., 2001; Stark and Okado, 2003).

The mechanisms of priming and familiarity are therefore related, in that both depend on synaptic changes between the semantic system and a perceptual system (differing mainly as a consequence of synaptic change from the semantic to the perceptual system in the case of priming, or from the perceptual to the semantic system in the case of familiarity). This is consistent with claims that priming and familiarity derive from the same cause—processing fluency (Jacoby and Dallas, 1981)—perhaps differing in practice according whether the task orients the participant toward perceptual fluency (prediction error in occipitotemporal cortex) or conceptual fluency (prediction error in perirhinal cortex). However the psychological concept of a "feeling of familiarity" is clearly more complex than such simple distinctions, since feelings of familiarity only arise when processing fluency is unexpected (e.g., one might process one's toothbrush fluently every evening, but this does not often lead to a feeling of familiarity attributed to the past, Whittlesea and Williams, 1998). This concept of familiarity should not therefore be directly equated with prediction error in perirhinal cortex; rather prediction error corresponds to fluency, whereas familiarity is the attribution of that fluency to the past. A full theory of the concept of familiarity is thus beyond current considerations, and needs to include meta-memory factors such as the expected level of fluency in a certain situation.

In the case of recollection, PIMMS states that a key function of the hippocampus is to optimize the mutual predictability between items (represented in perirhinal cortex) and contexts (presumably represented in multiple regions depending on the type of context, though here we will focus on spatial context, and assume this is represented in parahippocampal cortex). Mutual predictability corresponds to the joint probability of predicting an item from a context and predicting a context from an item. For example, even though a context may not strongly predict a particular item, an item may predict a different context, leading to high mutual prediction error in the hippocampus. It is this hippocampal prediction error that drives episodic encoding, i.e., synaptic change between hippocampus and perirhinal (and parahippocampal) cortices. Therefore those items that lead to high prediction error in hippocampus (for reasons including the precision of their item representation and the strength of their item-context associations; see "butcher in office" example below) will tend to produce greater hippocampus fMRI signal in the study phase of a recognition memory experiment that is associated with subsequent recollection in the test phase.

Later during the test phase, these study-induced synaptic changes mean that the re-presentation of that item (leading to a pattern of activity in perirhinal cortex) produces a better reinstantiation of a representation of its study context. However, because the current context of the recognition test phase does not match the reinstantiated context, this produces greater prediction error in the hippocampus (and possibly the parahippocampal cortex)—i.e., greater fMRI signal during recollection. Conversely, if no context is predicted by the item (i.e., no study context is reinstantiated), then there is little prediction error in the hippocampus to minimize, hence little fMRI signal

---

[iii]The dynamics of the resolution of prediction error may be visible with other techniques like EEG and MEG—indeed the parietal variant of the P600 that has been associated with recollection might relate to prediction error in the episodic system, while the earlier frontal variant of the N400 that has been associated with familiarity might relate to prediction error in the semantic system (see Rugg and Curran, 2007; Paller et al., 2007)—but such considerations are beyond the present remit.

[iv]Note that reduced activity to repeated stimuli can be found in other brain regions too, including the hippocampus, for example in single cell responses to scenes (Yanike et al., 2009).
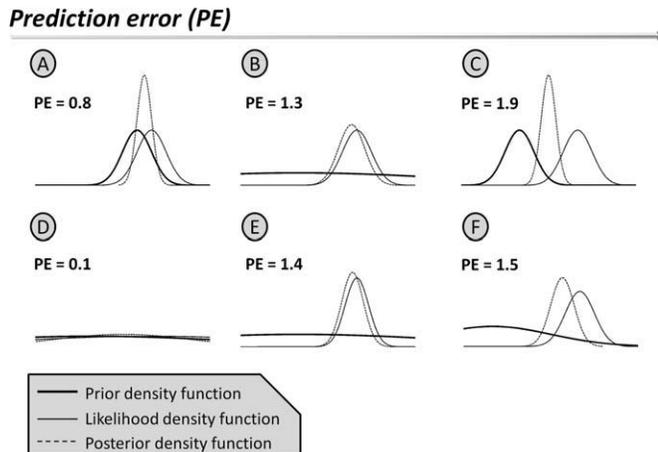
Prediction error (PE)

FIGURE 4. **Illustration of prior and likelihood probability density functions within a layer of topographically organized neurons comprising a memory system. The *x*-axis might reflect, for example, the features in a perceptual memory system, or items in a semantic memory system, organized by similarity along a single dimension for illustrative purposes (in reality the distributions might be multidimensional). The prior, from a higher layer in PIMMS, is shown in heavy lines; the data likelihood (or "evidence"), from a lower layer, is shown in light lines (and the posterior probability is shown by the dashed lines). The integrated difference between prior and evidence densities gives the initial prediction error (PE) shown in each panel. The top row shows cases with fixed evidence but different prior precisions; specifically: a precise and accurate prior (Panel A, e.g., "butcher in the butcher's shop"); an imprecise prior (Panel B, e.g., "butcher on the bus"); or a precise but inaccurate prior (Panel C, e.g., "butcher in one's office"). Panels D and E show the converse case of an imprecise (flat) prior together with either imprecise evidence (e.g., a meaningless stimulus) or more precise evidence (e.g., a primed stimulus) respectively. Panel F shows the case for low-frequency items, where the evidence is less precise, but the prior is biased toward other (high-frequency) items.**

there. In a simple toy example in the Supporting Information, we give a concrete example of how a simple two-layer version of PIMMS that distinguishes between these context-to-item associations and item-to-context associations can explain the qualitatively different patterns of fMRI signal in Figure 1C. Below, we expand the above accounts in terms of probability densities.

A Probabilistic Perspective on PIMMS

One way to think about PIMMS is in terms of prior and likelihood probability densities. For illustrative purposes, imagine a single layer of neurons with tuning curves that are topographically organized, such that similar contexts/items/features are represented close together. Probability densities reflecting interpretation of an input to that layer are then likely to be unimodal distributions (Fig. 4). The likelihood density (the light lines in Fig. 4) represents the data, or "evidence," coming from the system below, and the prior density (the heavy lines in Fig. 4) corresponds to the prediction coming from the sys-

tem above. The difference between the two density functions determines the initial prediction error. After settling into a steady state in which this prediction error is minimized, the activity in the semantic system, for example, can be interpreted as the posterior probability (the dashed lines in Fig. 4) of each item being present, given the contextual priors and perceptual evidence. The difference between the posterior and prior densities after the network has stabilized is then the residual prediction error that drives subsequent learning (such that the future prior densities are moved slightly toward the posterior density). Thus a high residual prediction error received by the semantic system from the perceptual system, for example, will lead to semantic encoding, while a high residual prediction error received by the episodic system from the semantic or perceptual systems will lead to episodic encoding.

The prior distributions in Figure 4 can have either low or high variance, corresponding to either precise ("strong") or imprecise ("weak") predictions, respectively. The evidence (likelihood) can be similarly precise or imprecise. In the extreme case where there are no predictions (for example in a completely novel context), the priors are "flat" (e.g., any item is deemed equally likely to occur). In addition, a prior can also be more or less accurate (depending on how close its central tendency is to the central tendency of evidence). Thus an "accurate" contextual prior means that an expected item is present, whereas an inaccurate prior means that an item different from expected is present. Assuming that the fMRI signal not only integrates neural activity over the time taken to minimize prediction error, but also integrates activity spatially across all neurons in the layer in Figure 4, then PIMMS predicts that the fMRI signal will tend to increase as the integrated difference (divergence) between the prior and likelihood densities increases.[v] We now show how considering priors and evidence in terms of both their precision and accuracy can help make testable predictions for everyday situations.

Everyday Memory and New Experimental Predictions From PIMMS

The importance of contextual priors: "The butcher in the office"

A probabilistic conception of PIMMS suggests some predictions for the fMRI signal in perirhinal cortex (semantic system) as a function of three different types of context: (1) a precise and accurate prior (Fig. 4A), (2) an imprecise prior (Fig. 4B), and (3) a precise but inaccurate prior (Fig. 4C). Because the prediction error increases across these three contexts (i.e., greater divergence between prior and likelihood distributions), the fMRI signal is expected to increase in perirhinal cortex.

This theoretical prediction may not be very surprising, but the assumption of PIMMS that the amount of subsequent synaptic change between perirhinal and hippocampal systems is determined by the amount of residual prediction error leads to the further experimental prediction that the occurrence of an item that is not expected in a context (Fig. 4C) will be better episodically encoded than occurrence of the same item in a context that has no strong expectancy (Fig. 4B).

An everyday example of this prediction would be a modification of the famous "butcher on the bus" example. This example is normally used to illustrate familiarity—i.e., a sense of knowing the person but not remembering who they are—but the question here is how well this event would be encoded (and in the examples below, we are assuming that the butcher is, in fact, recognized as the butcher). As a type of public transport, one could argue that the context of being on a bus does not impose strong predictions about the people one might encounter there. In the context of the butcher's shop however, there is a strong prior for the butcher, while in one's office at work, there are strong priors for one's work colleagues instead.[vi] According to the above argument, PIMMS predicts that one will encode the event of "butcher in the office" better than "butcher on the bus" (and both should of course be better encoded than the fully-expected "butcher in the butcher's shop"). In other words, if one later encountered the butcher in a new town (another context with weak priors); the prediction is that one would be more likely to recollect his previous appearance in one's office than his previous appearance on a bus. Testing the above prediction of PIMMS in the laboratory (i.e., fMRI scanner) requires imposing specific contexts and varying the precision of associations between those contexts and specific items. We are not aware of such an fMRI experiment to date, but would predict that such an experiment would show both increased perirhinal and hippocampal activity, and increased perirhinal-hippocampal connectivity (see below), for those conditions analogous to the "butcher in the office," relative to those analogous to the "butcher on the bus."

The importance of item priors: "The recently seen butcher"

The perspective of precision and accuracy can also help explain the role of prior item knowledge in recognition memory. Consider the case when the (episodic) prior on the semantic system is flat—i.e., no particular items are predicted in the current context—but the evidence from lower perceptual systems varies (Figs. 4D,E). If a completely novel stimulus is encountered (unlike any known item), then there will be a flat distribution of activity across neurons in the semantic system. This is because, even if there is reasonably precise input to the perceptual system, the lack of any unique predictions for that combination of features (in the item-to-feature and feature-to-item connections) results in an imprecise likelihood within the semantic system. The prediction error forwarded to the episodic system between a flat (episodic) prior and such flat evidence (Fig. 4D) is minimal (i.e., there is little information that can be learned). In turn, episodic encoding of such unknown items (in terms of weight changes between semantic and episodic systems) is minimal. This would be analogous to the standard "butcher on the bus" situation: if the butcher is not identified as such (the posterior distribution over the item layer remains imprecise), then that situation will not be well-encoded.[vii] If a stimulus is better known (or more closely resembles known items), on the other hand, there will be a more precise data likelihood, greater prediction error, and hence greater episodic encoding (Fig. 4E). This can explain the better recognition memory for novel stimuli (e.g., visual squiggles) that are found to be more "meaningful" (Voss and Paller, 2007).

This explanation may seem difficult to reconcile with evidence that recollection is often better for low than high frequency words (for review, Yonelinas, 2002), but in fact, PIMMS can predict greater episodic encoding for low than high frequency items if the contextual prior is assumed to be imprecise, but not flat—specifically, slightly biased toward high frequency items (Fig. 4F). While the evidence (input from perceptual systems) may be less precise for low- than for high-frequency items, the presence of some information in the feature-to-item associations from the perceptual layer (unlike the above case of a truly unknown stimulus) means that the data likelihood for a low frequency-item is not flat. Importantly, its central tendency will be further removed from that of the prior than would be the central tendency of a high-frequency item, which can lead to a greater relative prediction error for low-frequency items and hence a greater updating of item-context connections (i.e., greater episodic encoding; though effects of frequency at encoding are also likely to involve other factors such as attention; Diana and Reder, 2006).

Lastly, consider for example the effect of repetition priming on episodic encoding. Take an extended recognition memory

---

[vi]Note that these "semanticized" associations between a context (e.g., butcher's shop) and an item (e.g., butcher) may in fact be instantiated in direct connections between parahippocampal and perirhinal cortex, rather than via the hippocampus (an issue we will explore in extensions of PIMMS), but for the purpose of the predictions for the laboratory (fMRI) experiment below, we assume that experimentally-established context-item associations engage the hippocampus, i.e., the episodic system.

[vii]Note that we do not distinguish here between a representation of a familiar face, and a representation of the identity of that face (both are subsumed within the semantic system here). In a more complex model with yet further levels of representation, it might be possible to explain how a familiar face (the unrecognised butcher) can be well encoded (i.e., later recollected) as having occurred in a specific context (owing to plasticity between that layer and a contextual layer, e.g., hippocampus), even if that face is never identified as the butcher's. We are also aware that we do not address the issue of how new semantic representations of completely novel stimuli are ever learned, which is a difficult question that we leave for further developments of PIMMS, though is likely to entail changes in lateral connections within a region.

paradigm in which some of the items in the study phase are pre-exposed in a preceding "priming phase." One effect of this prior perceptual processing of a stimulus is to "sharpen" the feature-item connections, leading to a more precise distribution of evidential input into the semantic system. In conjunction with an uninformative contextual prior, this sharper likelihood produces greater prediction error and in turn greater episodic encoding (cf. Figs. 4B and E). Assuming that the context in most laboratory experiments conforms to such a weak prior (i.e., though a participant may become aware that only words or pictures will be presented, the probability of any particular exemplar of that category is normally fairly constant), this improved episodic encoding for perceptually primed relative to unprimed words is exactly what was found in a recent behavioral experiment (Gagnepain et al., 2008b; see also Poppenk et al., 2010).

Furthermore, the primed items are likely to have additionally been associated with a different context (that of the priming phase), which might be reinstated when they are presented again during the main study phase. According to PIMMS, this will in turn increase the mutual prediction error in the hippocampus, again affecting episodic encoding of the new study context. Returning to the above "butcher in the office" example, the strength of encoding in episodic memory will depend not only whether or not one has seen the butcher recently (i.e., whether his face has been primed), but also whether those occasions occurred in one or a varied set of contexts. This trade-off between effects at perceptual, semantic and episodic levels makes the precise patterns of local fMRI signal in each brain region difficult to predict uniquely. This is the kind of situation where the effective connectivity between brain regions might be better able to dissociate familiarity and recollection.

The importance of effective connectivity

As described above, PIMMS proposes that recollection and familiarity depend on complex spatial and temporal interactions between all three memory systems. Coupled with the temporal and spatial integration of the fMRI signal, this means that, while local patterns of fMRI signal within brain regions may be predicted in certain situations, it will be difficult to predict this pattern a priori (or uniquely) in many other situations. Nonetheless, another prediction made by PIMMS is that the pattern of "effective connectivity" between brain regions should differ during recollection and familiarity. "Effective connectivity" refers to changes in the functional coupling between regions as a consequence of an experimental manipulation.[viii] Increased functional coupling between two regions is

[viii]More precisely, effective connectivity refers to the causal influence that activity in one region has on that in another (Friston, 2002), as defined for example by an explicit network model, and can be framed dynamically in terms of how the rate of change of activity in one region depends on the level of activity in another.

generally expected when the prediction error in both regions is high (in that greater interaction between regions is needed to reduce that error), but can also occur when prediction error is rapidly reduced in one region because of activity in the other: i.e., effective connectivity from Region A to Region B is high when activity changes in Region B are highly dependent on activity in Region A. PIMMS predicts that familiarity will be associated with changes in the effective connectivity between semantic and perceptual systems, while recollection will be associated with increased effective connectivity between episodic and semantic and/or perceptual systems (Fig. 2B). Support for the latter hypothesis comes from a recent fMRI experiment that examined subsequent memory effects for primed and unprimed stimuli, using the extended recognition paradigm described above (Gagnepain et al., in press). In this experiment, the result of complex interactions between episodic and perceptual systems like those described above was that the neural correlates of encoding that led to subsequent recollection were not apparent in the local pattern of fMRI signal in the hippocampus per se, but rather in increased effective connectivity between the hippocampus and a brain region (in the lateral, superior temporal cortex) that was associated with the perceptual priming of the auditory words.

CONCLUSIONS

We have outlined a theoretical framework—PIMMS—and applied it to the topic of this special issue, viz. the neural correlates of recollection and familiarity. We have focused on evidence from fMRI on healthy human volunteers during recognition memory paradigms. However, we believe PIMMS can be extended to many other phenomena in memory (and perception), such as priming and recall paradigms, and data from methods such as lesion studies and single-cell recording, including data from brain regions outside the MTL.

While we have used the phrase "memory systems" to relate PIMMS to traditional theories of memory, we believe prediction-error-driven plasticity is a general property of brain regions, supporting both memory and perception within those same regions. Indeed, memory encoding (storage) and perception (memory retrieval) are intimately-related in terms of "off-line" vs. "on-line" reduction of prediction error, respectively. At the same time, we accept that a primary reason why different brain regions might support different types of memory is computational (in the sense of Marr, 1971): in that it seems advantageous, for example, to separate one system that stores what is unique about experiences (e.g., individuation/pattern separation)—an episodic system—from another whose primary function is to extract regularities across experiences (e.g., generalization/pattern completion, McClelland et al., 1995)—a semantic system. Our main point is that in most everyday situations, it also makes sense for these systems to operate together (e.g., to both individuate an event and adjust prior knowledge about the content of that event). This means that,

unless one or more memory systems are inactive (e.g., from a brain lesion), "on-line" or "performance" measures like behavioral outcomes or fMRI will reflect a complex interplay between each system.

Acknowledgments

The authors thank Tim Shallice and two anonymous reviewers for their helpful comments.

REFERENCES

Aggleton JP, Brown MW. 1999. Episodic memory, amnesia, and the hippocampal anterior thalamic axis. Behav Brain Sci 22:425–444.

Bar M. 2007. The proactive brain: Using analogies and associations to generate predictions. Trends Cogn Sci 11:280–289.

Berry CJ, Shanks DR, Henson RN. 2008. A unitary signal-detection model of implicit and explicit memory. Trends Cogn Sci 12:367–373.

Bowles B, Crupi C, Mirsattari SM, Pigott S, Parrent AG, Pruessner JC, Yonelinas AP, Kohler S. 2007. Impaired familiarity with preserved recollection after anterior temporal-lobe resection that spares hippocampus. Proc Nat Acad Sci USA 104:16382–16387.

Brown MW, Warburton EC, Aggleton JP. 2010. Recognition memory: Material, processes and subrtates. Hippocampus 20:1228–1244.

Brown MW, Xiang JZ. 1998. Recognition memory: Neuronal substrates of the judgment of prior occurrence. Prog Neurobiol 55:149–189.

Bubic A, von Cramon DY, Schubotz RI. 2010. Prediction, cognition and the brain. Front Hum Neurosci 4:1–15.

Buckner RL. 2010. The role of the hippocampus in prediction and imagination. Ann Rev Psychol 61:27–48.

Buckner RL, Wheeler ME, Sheridan MA. 2001. Encoding processes during retrieval tasks. J Cogn Neurosci 13:406–415.

Carr VA, Rissman J, Wagner AD. 2010. Imaging the human medial temporal lobe with high-resolution fMRI. Neuron 65:298–308.

Cowell RA, Bussey TJ, Saksida LM. 2010. Components of recognition memory: Dissociable cognitive processes or just differences in representational complexity? Hippocampus 20:1245–1262.

Davachi L. 2006. Item, context and relational episodic encoding in humans. Curr Opin Neurobiol 16:693–700.

Dayan P, Hinton GE. 1996. Varieties of Helmholtz Machine. Neural Networks 9:1385–1403.

Degonda N, Mondadori CR, Bosshardt S, Schmidt CF, Boesiger P, Nitsch RM, Hock C, Henke K. 2005. Implicit associative learning engages the hippocampus and interacts with explicit associative learning. Neuron 46:505–520.

Dehaene S, Changeux JP, Naccache L, Sackur J, Sergent C. 2006. Conscious, preconscious, and subliminal processing: A testable taxonomy. Trends Cogn Sci 10:204–211.

Diana RA, Reder LM. 2006. The low-frequency encoding disadvantage: Word-frequency affects processing demands. J Exp Psychol Learn Mem Cogn 32:805–815.

Diana RA, Yonelinas AP, Ranganath C. 2007. Imaging recollection and familiarity in the medial temporal lobe: A three-component model. Trends Cogn Sci 11:379–386.

Dunn JC, Kirsner K. 1988. Discovering functionally independent mental processes: The principle of reversed association. Psychol Rev 95:91–101.

Eichenbaum H, Yonelinas AP, Ranganath C. 2007. The medial temporal lobe and recognition memory. Annu Rev Neurosci 30:123–152.

Eustache F, Desgranges B. 2008. MNESIS: Towards the integration of current multisystem models of memory. Neuropsychol Rev 18:53–69.

Fernandez G, Tendolkar I. 2006. The rhinal cortex: "Gatekeeper" of the declarative memory system. Trends Cogn Sci 10:358–362.

Friston KJ. 2002. Functional integration and inference in the brain. Prog Neurobiol 68:113–143.

Friston KJ. 2010. The free-energy principle: A unified brain theory? Nat Rev Neurosci 11:127–138.

Gagnepain P, Chetelat G, Landeau B, Dayan J, Eustache F, Lebreton K. 2008a. Spoken word memory traces within the human auditory cortex revealed by repetition priming and functional magnetic resonance imaging. J Neurosci 28:5281–5289.

Gagnepain P, Lebreton K, Desgranges B, Eustache F. 2008b. Perceptual priming enhances the creation of new episodic memories. Consc Cogn 17:276–287.

Gagnepain P, Henson R, Chetelat G, Desgranges B, Lebreton K, Eustache F. Is neocortical-hippocampal connectivity a better predicator of subsequent recollection than local increases in hippocampal activity? New insights on the role of priming. J Cogn Neurosci (in press).

Graf P, Schacter DL. 1985. Implicit and explicit memory for new associations in normal and amnesic subjects. J Exp Psychol Learn Mem Cogn 11:501–518.

Graham KS, Simons JS, Pratt KH, Patterson K, Hodges JR. 2000. Insights from semantic dementia on the relationship between episodic and semantic memory. Neuropsychologia 38:313–324.

Green DM, Swets JA. 1966. Signal Detection Theory and Psychophysics. New York: Robert E. Kreiger.

Greve A, Donaldson DI, van Rossum MC. 2010. A single-trace dual-process model of episodic memory: A novel computational account of familiarity and recollection. Hippocampus 20:235–251.

Hayman CAG, Tulving E. 1989. Is priming in fragment completion based on a "traceless" memory system? J Exp Psychol Learn Mem Cogn 15:941–956.

Henson RN. 2003. Neuroimaging studies of priming. Prog Neurobiol 70:53–81.

Henson RN. 2005. What can functional imaging tell the experimental psychologist? Q J Exp Psychol A 58:193–233.

Henson RN. 2006. Forward inference in functional neuroimaging: Dissociations vs. associations. Trends Cogn Sci 10:64–69.

Henson RN, Price CJ, Rugg MD, Turner R, Friston KJ. 2002. Detecting latency differences in event-related BOLD responses: Application to words versus nonwords and initial versus repeated face presentations. Neuroimage 15:83–97.

Henson RN, Cansino S, Herron JE, Robb WG, Rugg MD. 2003. A familiarity signal in human anterior medial temporal cortex? Hippocampus 13:301–304.

Humphreys MS, Pike R, Bain JD, Tehan G. 1989. Global matching: A comparison of SAM, Minerva II, Matrix and TODAM models. J Math Psych 33:36–67.

Jacoby LL. 1992. A process dissociation framework: Separating automatic from intentional uses of memory. J Mem Lang 30:513–541.

Jacoby LL, Dallas M. 1981. On the relationship between autobiographical memory and perceptual learning. J Exp Psychol Gen 110:306–340.

Johnson MK, Hashtroudi S, Lindsay DS. 1993. Source monitoring. Psychol Rev 114:3–28.

Mandler G. 1980. Recognizing: The judgment pf previous occurence. Psychol Rev 87:252–271.

Marr D. 1971. Simple memory: A theory for archicortex. Phil Trans R Soc Lond 262:23–81.

McClelland JL, McNaughton BL, O'Reilly RC. 1995. 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.

Mishkin M, Varga-Khadem F, Gadian D. 1998. Amnesia and the organization of the hippocampal system. Hippocampus 8:212–216.

Mumford D. 1992. On the computational architecture of the neocortex. II. The role of cortico-cortical loops. Biol Cybern 66:241–251.

Murray EA, Bussey TJ, Saksida LM. 2007. Visual perception and memory: A new view of medial temporal lobe function in primates and rodents. Annu Rev Neurosci 30:99–122.

Norman KA. 2010. How hippocampus and cortex contribute to recognition memory: Revisiting the complementary learning systems model. Hippocampus 20:1217–1227.

Norman KA, O'Reilly RC. 2003. Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. Psychol Rev 110:611–646.

Paller KA, Wagner AD. 2002. Observing the transformation of experience into memory. Trends Cogn Sci 6:93–102.

Paller KA, Voss JL, Boehm SG. 2007. Validating neural correlates of familiarity. Trends Cogn Sci 6:243–250.

Patterson K, Nestor PJ, Rogers TT. 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. Nat Rev Neurosci 12:976–987.

Poldrack RA. 1996. On testing for stochastic dissociations. Psychon Bull Rev 3:434–448.

Poppenk J, McIntosh AR, Craik FI, Moscovitch M. 2010. Past experience modulates the neural mechanisms of episodic memory formation. J Neurosci 30:4707–4716.

Ranganath C. 2010. A unified framework for the functional organization of the medial temporal lobes and the phenomenology of episodic memory. Hippocampus 20:1263–1290.

Rao RP, Ballard DH. 1999. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87.

Richardson-Klavehn A, Clarke AJB, Gardiner JM. 1999. Conjoint dissociations reveal involuntary "perceptual" priming from generating at study. Consc Cogn 8:271–284.

Rotello CM, Macmillan NA, Reeder JA. 2004. Sum-difference theory of remembering and knowing: A two-dimensional signal-detection model. Psychol Rev 111:588–616.

Rugg MD, Curran T. 2007. Event-related potentials and recognition memory. Trends Cogn Sci 11:251–257.

Schacter DL, Tulving E. 1994. Memory Systems 1994. Cambridge: MIT Press.

Shimamura AP. 2010. Heirarhical relational binding in the medial temporal lobe: The strong get stronger. Hippocampus 20:1206–1216.

Slotnick SD. 2010. Does the hippocampus mediate objective binding or subjective remembering? Neuroimage 49:1769–1776.

Squire LR, Wixted JT, Clark RE. 2007. Recognition memory and the medial temporal lobe: A new perspective. Nat Rev Neurosci 8:872–883.

Staresina BP, Davachi L. 2008. Selective and shared contributions of the hippocampus and perirhinal cortex to episodic item and associative encoding. J Cogn Neurosci 20:1478–1489.

Stark CE, Okado Y. 2003. Making memories without trying: Medial temporal lobe activity associated with incidental memory formation during recognition. J Neurosci 23:6748–6753.

Tulving E. 1985. Memory and consciousness. Can Psychol 26:1–12.

Tulving E. 2002. Episodic memory: From mind to brain. Annu Rev Psychol 53:1–25.

Tulving E, Gazzaniga MS. 1995. Organization of Memory: Quo Vadis? The Cognitive Neurosciences. Cambridge: MIT Press. pp 839–847.

Voss JL, Paller KA. 2007. Neural correlates of conceptual implicit memory and their contamination of putative neural correlates of explicit memory. Learn Mem 14:259–267.

Wais PE, Mickes L, Wixted JT. 2008. Remember/know judgments probe degrees of recollection J Cogn Neurosci 20:400–405.

Whittlesea BWA, Williams LD. 1998. Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. Acta Psychol 98:141–165.

Wixted JT, Mickes L, Squire LR. 2010. Measuring recollection and familiarity in the medial temporal lobe. Hippocampus 20:1195–1205.

Yanike M, Wirth S, Smith AC, Brown EN, Suzuki WA. 2009. Comparison of associative learning-related signals in the macaque perirhinal cortex and hippocampus. Cereb Cortex 19:1064–1078.

Yonelinas AP. 2002. The nature of recollection and familiarity: A review of 30 years of research. J Mem Lang 46:441–517.

Yonelinas AP, Aly M, Wang W-C, Koen JD. 2010. Recollection and familiarity: Examining controversial assumptions and new directions. Hippocampus 20:1178–1194.