# Chapter 1

# Introduction

Learning and memory, the ability to acquire, store and retrieve information, shapes the way we interact with our environment. This thesis describes the functional properties of the hippocampus, a neuroanatomical region critical for a particular type of memory - long-term memory that is accessible to conscious recollection. The history of human neuropsychological and animal studies leading to the discovery of this specific hippocampal role in memory is presented. Despite the precise definition of the hippocampal mnemonic role, the operations carried out by the hippocampus during sensory processing, to enable learning and memory, is still unresolved.

An efficient memory system requires the ability to detect and encode novel stimuli. In this thesis it will be argued, on the basis of functional magnetic resonance imaging (fMRI) experiments in normal human subjects, that the functions of anterior (rostral) and posterior (caudal) hippocampus dissociate. It will be suggested that anterior hippocampus is engaged by mismatches between expectation and experience, a function that may be central for allowing sensory input access to long-term memory. By contrast, evidence will be presented for a posterior hippocampal role in processing familiar stimuli. Hence this region of hippocampus may mediate aspects of retrieval from long-term memory.

## 1.1 Anatomy

The human hippocampus is located in the medial temporal lobe, lying on the floor of the inferior horn of the lateral ventricle. The hippocampus proper is composed of the CA (cornu ammonis) subfields whereas the hippocampal formation includes the dentate gyrus (DG), CA subfields and subiculum. The term hippocampus will be used here to refer to all three components. All hippocampal components are composed of simple three-layered allocortex, which differentiates them from surrounding six-layered medial temporal neocortex, the entorhinal, perirhinal and parahippocampal cortices (Witter *et al*., 1989b).

The most anterior part of the human hippocampus begins rostromedially in the uncus of the medial temporal lobe. Posteriorly, the hippocampus extends into the atrium of the lateral ventricle where the efferent fibres of the fornix ascend around the posterolateral aspect of the thalamus. In the anteroposterior direction, the human hippocampus is approximately 30 mm long but it is curved in the dorso-ventral plane, making the longitudinal axis roughly 50 mm in length (Rosene and Van Hoesen, 1987).

The hippocampus is traditionally considered part of the so-called limbic system. In 1878, Broca categorised the subcallosal, cingulate and parahippocampal gyri, as well as the hippocampus, as the limbic lobe. Papez (1937) related emotional disturbances experienced by some of his patients to hippocampal and cingulate damage. He proposed the existence of a closed circuit important for elaborating and expressing emotion. This ‘Papez circuit’ thus became defined as a sequence of

interconnected structures. In this circuit, the hippocampus projects by way of the fornix to the mammillary bodies, which in turn project to anterior nuclei of the thalamus. From the thalamus, the cingulate cortex is reached, and finally there is a projection back to the hippocampus via the parahippocampal region. MacLean (1952) introduced the term limbic system to place both limbic cortex and associated subcortical structures (mammillary bodies, anterior thalamus and the amygdala) into one functional system. The entire concept of a unified limbic system has now been challenged (Brodal, 1981). However, although this thesis focuses on the hippocampus, the above considerations serve to illustrate broader contexts in which the hippocampus may exert its functional role.

The major projection to the hippocampus, the perforant path, originates in surrounding entorhinal cortex (Ramon y Cajal, 1911). The entorhinal cortex, and other components of medial temporal cortex, namely perirhinal and parahippocampal cortices, are reciprocally connected with widespread areas of neocortex, particularly association cortex (Van Hoesen, 1982). Based on detailed studies of cortical connectivity, it has been proposed that within the cortex, information from all sensory modalities is processed through a sequence of projections. During transfer from primary areas, via secondary and tertiary unimodal association areas, towards multimodal association areas, information becomes more elaborated or complex (Pandya and Seltzer, 1982; Van Essen and Maunsell, 1983). Since the medial temporal cortical regions are strongly interconnected with most multimodal areas (Van Hoesen, 1982), this cortical area can be viewed as supramodal cortex where all cortical channels converge. This convergence is taken one step further in the hippocampus, where interactions with subcortical and brainstem projections occur

(see Witter *et al*., 1989b for review). Hippocampal connectivity is therefore particularly suited to a mnemonic role in that it receives information not only about the external world through sensory inputs but also interoceptive information, from subcortical and brainstem systems, regarding the internal state of the organism.

### 1.2 Discovery of multiple memory systems in the brain

Learning how to ride a bike seems, intuitively, a different process than learning that Paris is the capital of France. In 1949, the philosopher of mind, Gilbert Ryle, proposed the existence of two types of knowledge: knowing how, e.g. knowing how to ride a bike, and knowing that, e.g. knowledge of facts. Despite this dissociation in knowledge types, it had not yet been conceived that different brain regions could support different types of memory. In Ryle's time the prevailing theory of brain function was Lashley's law of mass action, according to which the extent of memory defect subsequent to brain damage was dependent upon the size of the cortical area lesioned, not with its specific location (Lashley, 1929). Although ideas of functional specialisation were acknowledged since the late 19$^{th}$ century, following the observations of Broca (1861) and Wernicke (1874), it was not until the 1950s that a specific memory function was localised to a distinct neuroanatomical region.

In 1957, Scoville and Milner described the famous patient HM, who had sustained bilateral resection of the medial temporal lobe to relieve severe epilepsy. It was immediately evident following surgery that HM had a very profound impairment of recent memory in the apparent absence of other intellectual loss. He could not remember what he had for breakfast, and he could not find his way around the

hospital or recognise members of hospital staff. HM was unable to acquire any new information that could be later recalled, although he could hold immediate impressions in his mind provided interfering activity did not distract his attention. In contrast, old memories from his childhood seemed to be intact (Scoville and Milner, 1957).

HM was able to register perceptual information normally, but this information ceased to be available to him after 30-40 seconds (Milner, 1972). On a test of delayed pair comparison, where two stimuli are presented in succession, separated by a short time interval, normal subjects are able to indicate whether the second stimulus is the same as the first, making very few errors with a 60 second delay and interpolated distraction. HM performed accurately at zero delay, but with increasing intratrial intervals, his performance decayed rapidly; 60 second delay scores approached chance levels and were not further impaired by distraction (Milner, 1972). An identical finding was reported by Sidman *et al*. (1968), using a delayed matching-to-sample technique, where subjects must chose, after a variable intratrial interval, which of a number of stimuli was presented previously as the sample. HM was at chance after a 32 second delay.

Further neuropsychological characterisation of HM revealed striking dissociations in the memory deficit arising from bilateral medial temporal removal. Despite HM's deficit in the above tasks, bilateral medial temporal lobectomy did not impair the acquisition of new motor skills (Milner, 1962). Like normal subjects, HM could, across trials, improve his ability to draw the outline of a star in a mirror. HM learned this new skill despite having no recollection, at the end of learning, of ever

having performed the mirror drawing task. This finding provided early evidence for dissociable memory systems in the brain.

*1.21 Declarative and nondeclarative memory systems*

Further work demonstrated that motor skills are a subset of the learning and memory abilities that are spared following bilateral medial temporal lobe lesions. The dissociations in memory deficits following bilateral medial temporal lesions in HM and other amnesic patients led to a classification of memory types based on the dependence on medial temporal integrity. This classification addresses long-term memory, which is different from short-term memory, such as digit-span, which is intact following medial temporal damage (Shallice and Warrington, 1970).

The major distinction between different types of long-term memory is between declarative (or explicit) memory and a collection of nondeclarative (or implicit), nonconscious forms of memory (Cohen and Squire, 1980). Declarative memory is dependent on the medial temporal lobes and affords the capacity for conscious recollection of episodes and facts (Squire, 1992). Declarative memory is propositional, being either true or false. It is fast, not always reliable (i.e. forgetting and retrieval failure can occur), and flexible in the sense that declarative memories are accessible to multiple response systems. Declarative memory can be further divided into episodic memory, memory for events that compose a unique personal experience, and semantic memory, factual information that is independent of the specific episodes in which that information was acquired (Tulving, 1972). These two components of declarative memory can be shown to dissociate if one tries to remember the episode that led to the learning of a particular fact. We know that Paris

is the capital of France but are unlikely to remember the episode that led to the learning of this fact.

Nondeclarative memory is neither true nor false. In the case of nondeclarative memory, performance changes as a result of experience, which justifies the term memory, but performance changes without providing conscious access to any prior episodes (Schacter and Tulving, 1994). This type of memory is fully intact in the presence of hippocampal damage (Squire *et al*., 1993). It underlies changes in skilled behaviour, such as mirror drawing described above, and the ability to respond appropriately to stimuli through practice, as a result of conditioning or habit learning. It also includes changes in the ability to process objects as a result of recent encounters, a phenomenon known as priming (Warrington and Weiskrantz, 1968). Nondeclarative memory is slow in acquisition (priming is an exception), reliable, and inflexible in that information is not readily expressed by response systems not involved in the original learning.

Priming is perhaps the best studied form of nondeclarative memory in humans. Priming was first detailed by Warrington and Weiskrantz (1968) who asked subjects to identify line drawings of common objects and animals from which most of the contour lines had been removed. Over successive presentations, the contour was gradually filled in until subjects could name the item. On a second presentation of the task, 1 hour later, normal subjects showed considerable improvements in performance, requiring fewer contour cues to name the items. On this incomplete figures task and on an analogous fragmented words task, Warrington and Weiskrantz (1968) found marked priming in a right medial temporal lobectomy patient, with

good retention 4 weeks later, despite this patient not remembering doing the task before (i.e. the patient was amnesic).

Further investigations of priming have used paradigms in which subjects see a list of words, pictures of objects, or non-verbal material such as novel objects or designs. Subsequently, subjects are tested with both old and new items and asked to name words or objects as quickly as possible, to complete fragments to form whole items, or to make rapid decisions about items. For example, when the first few letters MOT__ of a recently studied word MOTEL are presented, priming is evident in the tendency to complete the word fragment to form the studied word instead of other possible words. Patients with medial temporal damage exhibit fully intact priming, despite being unable to recognise as familiar the items that had been presented previously (Tulving and Schacter, 1990). These forms of perceptual priming are thought to be mediated by higher visual cortical areas, such as inferior temporal regions (Schacter and Bucker, 1998). Priming is, however, not limited to the perceptual domain. Various forms of conceptual priming also exist (Roediger and MacDermott, 1993). Conceptual priming reflects enhanced task performance subsequent to prior processing of stimulus meaning, such as category-exemplar generation.

In mammals, an intensively studied example of nondeclarative memory is classical Pavlovian conditioning of discrete behavioural responses. A conditioned response develops when a previously neutral stimulus (CS) is consistently and unavoidably followed by another stimulus (US). The US automatically elicits a behavioural response (UR) because of, for example, its aversive nature (such as a

shock) or appetitive value. After a CS is repeatedly paired with a US, the CS, when presented alone, elicits a conditioned response (CR) identical to that produced by the US. In the rabbit eyeblink response, a widely studied model of conditioning (Thompson and Krupa, 1994), the CS is typically a tone, the US an airpuff to the eye and the CR an eyeblink to the tone. Based on anatomical findings, electrical stimulation, and reversible lesion techniques, there is strong evidence that the essential memory trace circuit includes the cerebellum and related brain stem circuitry and that the memory traces themselves are formed and stored in the cerebellum (Thompson and Krupa, 1994). This simple form of learning is, therefore, not dependent on the hippocampus, supported by the fact that patients with medial temporal damage show intact Pavlovian eyeblink conditioning (Weiskrantz and Warrington, 1979). Interestingly, trace conditioning, in which a temporal delay is introduced between the CS and delivery of the US, is thought to be hippocampus-dependent (Clark and Squire, 1998).

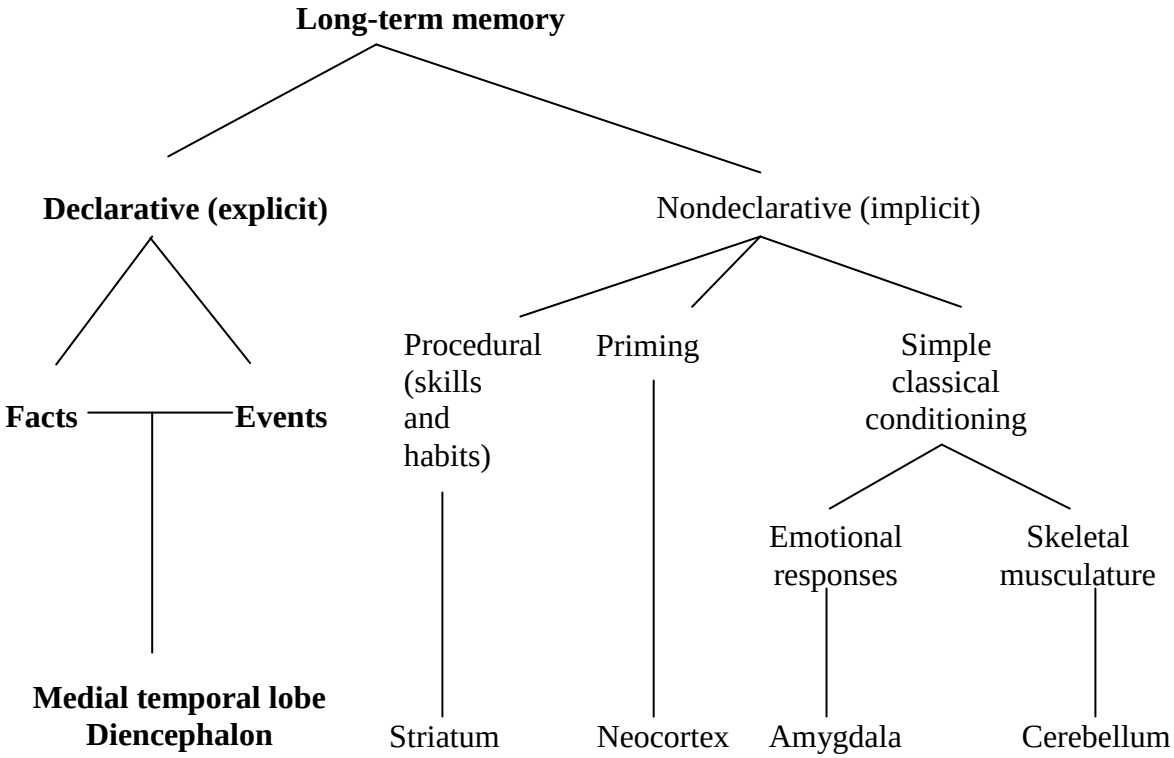
Further forms of nondeclarative memory include habit formation, which refers to the gradual acquisition of tendencies or dispositions that are specific to a set of stimuli that guide behaviour. Habit learning survives hippocampal damage in humans and experimental animals but is impaired by damage to the caudate nucleus of the basal ganglia (Packard *et al*., 1989; Knowlton *et al*., 1996).

Particular aspects of emotional learning, such as the development of phobias or fear conditioning, are dependent on the amygdala and are intact following hippocampal lesions (Aggleton, 1992; Adophs *et al*., 1997). Knowledge about the essential structures and connections involved in emotional learning has come from

extensive research into fear conditioning and fear-potentiated startle arising when rats learn to fear a neutral stimulus such as a tone (LeDoux, 1995). Integrity of the hippocampus is, however, thought to be required for the rat to be conditioned to a particular context (contextual fear conditioning; see Holland and Bouton, 1999 for review).

The amygdala has also been shown to be important for emotional learning in humans (Bechara *et al.*, 1995; Damasio, 1995; Cahill *et al*., 1996; Buchel *et al*., 1998) and this role will be discussed in chapter 5. Amygdala damage does not, however, produce impairments in declarative memory (Bechara *et al*., 1995), providing a double dissociation between amygdala and hippocampal roles in emotional and declarative learning, respectively. Also described in chapter 5 is the fact that in addition to the essential role for the amygdala in emotional learning, the amygdala is thought to exert modulatory effects on other memory systems (Cahill and McGaugh, 1998). Specifically, it has been suggested that the amygdala mediates the enhancement of declarative, conscious memories evoked by emotional arousal (Chapter 5; Adolphs *et al*., 1997) by augmenting hippocampal processing (Cahill and McGaugh, 1998).

**Figure 1.1. Classification of long-term memory.**



### 1.3 The Amnesic syndrome

A classification of memory types, described above, is summarised in figure 1.1. This classification has enabled a precise definition of the memory deficit arising from medial temporal damage. Damage to the medial temporal lobes produces the amnesic syndrome, a deficit in declarative memory (Squire and Zola-Morgan, 1991). The amnesic syndrome is characterised by a dense anterograde amnesia, an inability to acquire any new declarative information that is later accessible to conscious recollection, plus a variable, graded retrograde amnesia. The retrograde amnesia is described as variable because the extent of retrograde amnesia differs between patients (Nadel and Moscovitch, 1998). It is described as graded because recent declarative memories are most severely affected, whilst memories acquired long ago, such as childhood memories, tend to be spared (Scoville and Milner, 1957; Squire and Alvarez, 1995).

The graded nature of retrograde amnesia has led to a suggestion that the hippocampal role in memory is time-limited (Zola-Morgan and Squire, 1990). The classical model holds that initially, the hippocampus and related structures are necessary for the storage and recovery of memory traces, but their contribution diminishes as consolidation proceeds, until the neocortex alone is capable of sustaining the permanent memory trace and mediating its retrieval (Squire and Zola-Morgan, 1991). There have, however, been a number of reports suggesting that the gradient of retrograde amnesia following medial temporal damage is flat i.e. remote declarative memories are not selectively spared (Barr *et al.*, 1990; Warrington and Duchen, 1992; Cippolotti *et al*., 2001; see also Nadel and Moscovitch, 1997).

Although damage limited to the hippocampus proper in humans is sufficient to produce the amnesic syndrome (Zola-Morgan *et al*., 1986), both the anterograde and retrograde components of the amnesic syndrome are more severe following more extensive medial temporal damage including surrounding cortical areas (Rempel-Clower *et al*., 1996; Reed and Squire, 1998). Nadel and Moscovitch (1997) have suggested that the graded nature of retrograde amnesia can be explained by the extent of medial temporal damage. Specifically, they argue that the temporal extent of retrograde amnesia is proportional to the extent of medial temporal damage. In summary, it is widely accepted that hippocampal lesions produce anterograde memory impairment. The suggestion that the hippocampal role in memory is time-limited (Zola-Morgan and Squire, 1990), remains a matter of on-going debate.

*1.31 Semantic memory and the hippocampus*

The taxonomy of memory presented in figure 1.1 states that both event and factual memory comprise medial temporal-dependent declarative memory. While it is agreed that episodic memory is severely impaired in amnesia and dependent on the medial temporal memory system, the relationship between semantic memory and this system is unclear (Mishkin *et al*., 1998). Despite the variable retrograde amnesia for events, patients with the amnesic syndrome have relatively preserved general knowledge (Mayes, 1988). Amnesic patients do have difficulty acquiring semantic knowledge (Glisky *et al*., 1986; Kovner *et al*., 1983), but they can typically succeed, to variable degrees, with repetition. For example, the severely amnesic patient KC eventually learned to complete arbitrary three-word sentences during a large number of training trials distributed over several months (Tulving *et al*., 1991). This occurred

despite an apparent absence of conscious memory for specific episodes. Similarly, Kitchener *et al*., (1998) reported an amnesic patient who acquired new vocabulary and information about famous people and public events in the 13-year period following a left medial temporal infarct, despite having no measurable anterograde episodic memory and a profound loss of autobiographical memory. In children with selective hippocampal damage and episodic memory impairments, acquisition of semantic memory appears intact (Vargha-Khadem *et al*., 1997). However, despite these observations, the role of the hippocampus in semantic memory remains controversial. Mishkin and colleagues have recently suggested that episodic and semantic memory may be mediated by different components of the medial temporal memory system, with hippocampus supporting episodic memory and medial temporal cortices supporting semantic memory (Mishkin *et al*., 1998). This suggestion has yet to be demonstrated empirically.

*1.32 Critical components of the medial temporal lobe memory system*

The results of a recent magnetic resonance imaging study of HM's medial temporal lesions (Corkin *et al*., 1997) confirmed that Scoville's bilateral resection had included the amygdala, the entorhinal and perirhinal cortices, and the anterior hippocampus. On the other hand, the parahippocampal cortex was largely spared, along with posterior hippocampus. The anatomical imprecision of human clinical evidence has led researchers to model anterograde amnesia in animals, allowing the testing of more selective lesions. Animals do not express their memories by verbal declaration, and whether they have the capacity for conscious recollection is a matter of debate. However animal investigations into hippocampal function have provided

valuable insight into the precise functional roles of the components of medial temporal lobe.

Work with monkeys, together with recent, more detailed anatomical information from amnesic patients, has identified the structures within the medial temporal lobe that are critical for memory. These structures are the hippocampus and adjacent cortical areas that are anatomically related to the hippocampus, namely the entorhinal, perirhinal and parahippocampal cortices (Squire and Zola Morgan, 1991). Monkeys with lesions to the medial temporal lobes are, like human amnesic patients, impaired on tasks of declarative memory. As in human amnesia, the impairment is multimodal (Murray and Mishkin, 1984) while skill and habit learning and other tasks of a nondeclarative type are intact (Zola-Morgan and Squire, 1984; Zola-Morgan and Squire, 1993). Immediate memory (i.e. short-term memory) is also spared (Overman *et al*., 1990).

Work with monkeys has involved several tasks known to be sensitive to human amnesia (Aggleton *et al*., 1988; Squire *et al*., 1988), including retention of simple object discriminations and the simultaneous learning of multiple pairs of objects (e.g. eight-pair concurrent discrimination learning). The most widely used task has been trial-unique delayed non-matching to sample (Mishkin and Delacour, 1975). In this test of recognition memory, the monkey first sees a sample object then, after a delay, the original sample object and a novel object are presented together. The monkey must displace the novel object to obtain a food reward. New object pairs are used in each trial.

Studies in the monkey started with a large medial temporal lobe resection that approximated that of HM (Mishkin, 1978). This lesion was termed the $H^+A^+$ lesion (Squire and Zola-Morgan, 1988) where H refers to the hippocampus; A, the amygdala; and $^+$, the cortical regions adjacent to the hippocampus and amygdala that are necessarily damaged when either of these structures are removed using a direct surgical approach (i.e. the perirhinal, entorhinal, and parahippocampal cortices). The $H^+A^+$ lesion produces severe memory impairment (Mishkin, 1978; Mahut *et al*., 1981). Memory is also impaired following a lesion that involves only the posterior portion of the medial temporal lobe (the $H^+$ lesion), although the impairment is not as severe as with the $H^+A^+$ lesion (Mishkin, 1978). The $H^+$ lesion involves the hippocampus proper, dentate gyrus and subiculum, the posterior portion of entorhinal cortex, and parahippocampal cortex.

Later studies indicated that the more severe memory impairment associated with the $H^+A^+$ lesion, as compared with $H^+$ lesion, resulted from cortical, not from amygdala, damage. Monkeys with damage restricted to the amygdala (the A lesion), that spares underlying cortex, perform as well as normal monkeys on different memory tasks, including delayed non-matching to sample (Zola-Morgan *et al*., 1989a). If the $H^+$ lesion is extended forward to include the amygdala (the $H^+A$ lesion), the memory impairment associated with the $H^+$ lesion is not exacerbated (Zola-Morgan *et al*., 1989a). Monkeys with lesions of perirhinal and parahippocampal cortices (the PRPH lesion), which includes damage to projections to entorhinal cortex from other cortical areas, exhibit long lasting memory impairment in both visual and tactile modalities (Zola-Morgan *et al*., 1989b; Suzuki

*et al*., 1993). Similar findings have been reported in several studies involving rodents (see Squire, 1992 for review).

These findings made it apparent that the cortex adjacent to the amygdala, the perirhinal and entorhinal cortices, were of critical importance for declarative memory function. Neuroanatomical evidence demonstrates that the perirhinal and caudally adjacent parahippocampal cortices provide nearly two thirds of the cortical input to the entorhinal cortex (Insausti *et al*., 1987). Because entorhinal cortex is, in turn, the major source of projections to the hippocampus, there was reason to suspect that perirhinal damage would cause memory impairment. Indeed, if the $H^{+}$ lesion is extended forward to include the perirhinal cortex (the $H^{++}$ lesion), memory impairment is greater than after the $H^{+}$ lesion, which already has parahippocampal damage (Zola-Morgan *et al*., 1993).

The perirhinal, entorhinal and parahippocampal cortices are not merely routes by which information from the neocortex reaches the hippocampus. The fact that the memory deficit subsequent to hippocampal damage is more severe when these cortical regions are damaged (e.g. $H^{++}$ lesion versus $H^{+}A$) indicates that these cortical regions must also be important by themselves for memory function. The implication is that information from neocortex need not actually reach the hippocampus itself for some memory storage to occur. The role of perirhinal cortex in human memory will be further discussed in chapter 6. The hippocampus remains essential for declarative memory, as damage limited to the hippocampal region produces long-lasting memory impairment in both humans (Zola-Morgan *et al*., 1986) and monkeys (Alvarez *et al*., 1995).

*1.33 Hippocampal role in recall vs recognition*

Retrieval from episodic memory can be expressed either through recall or recognition. Early studies on HM demonstrated impairments in both recall and recognition tasks. Investigations of selective hippocampal lesions have, however, generated controversy regarding the hippocampal contribution to recognition processes. Loss of recognition memory was initially considered a core deficit of anterograde amnesia (Squire and Knowlton, 1995). A study evaluating recall and recognition performance of amnesic patients over a wide range of retention intervals found that recognition was impaired proportionately to recall (Haist *et al*., 1992). However, this claim was challenged by the findings from a recent survey of amnesics (Aggleton and Shaw, 1996) which analysed results from the Warrington Recognition Memory Test (RMT), a standard test of recognition. The RMT (Warrington, 1984) tests both face and word recognition. From a sample of 112 amnesics categorised into 11 different pathological groupings it appeared that patients with either restricted hippocampal damage (following hypoxia), or patients with fornix damage, performed as well on the RMT as age matched controls and performed significantly better than other amnesic groups (Aggleton and Shaw, 1996). The results of this survey are in accord with a recent report that patients with early, selective hippocampal damage are largely unimpaired in recognition memory (Vargha-Khadem *et al*., 1997).

In direct contradiction to these findings, a recent review of recognition performance following anoxic hippocampal damage in humans (Reed and Squire, 1997) demonstrated recognition deficits across a wide range of tests. However, compared to controls, performance on the RMT was apparently preserved in some patients and impaired in others (Reed and Squire, 1997). This discrepancy may be

due to variable covert pathology arising from anoxia, a limitation in interpreting results from patients with this aetiology.

The above debate in the human clinical literature regarding the role of human hippocampus in recognition processes is mirrored by findings in monkeys. The delayed nonmatching to sample task in monkeys is a standard test of recognition memory. It is a matter of controversy whether isolated lesions of the hippocampus produce impairments on this task (Alvarez *et al*., 1995; but see Murray and Mishkin, 1996; Aggleton and Brown, 1999). Lesion studies presented in the previous section (section 1.24) provide strong evidence that monkey perirhinal cortex is critical for recognition memory performance. In light of inconclusive evidence for a hippocampal role in recognition, a recent proposal is that hippocampal and perirhinal contributions to declarative memory dissociate, with hippocampus engaged by processes enabling conscious recall while perirhinal cortex is critical for recognition memory (Aggleton and Brown, 1999).

Declarative memory has also been shown to depend on the integrity of another component of the limbic system, the medial diencephalon, the critical structures including anterior and medial thalamic nuclei and mamillary bodies (Aggleton and Brown, 1999). Medial diencephalic damage, such as that observed in Korsakoff's psychosis, gives rise to a pattern of memory deficits identical to that produced by medial temporal damage (Victor *et al*., 1971), hence in many studies of amnesia, the two patient groups are pooled together.

*1.34 Hippocampal role in memory – current consensus*

From the preceding discussion, it follows that it is now widely accepted that hippocampal damage impairs episodic memory. The debate regarding the nature of the episodic deficit of the amnesic syndrome (i.e. time-limited role, recall vs recognition) has, however, prevented a consensus on the precise hippocampal functional role in episodic memory encoding, storage or retrieval. Furthermore, it is not known what operation the hippocampus carries out on incoming sensory information that enables it to be incorporated into episodic memory. Hippocampal lesions produce an anterograde amnesia. This deficit in acquiring new episodic memories suggests that hippocampal lesions may preferentially impair the processing of novel stimuli.

## 1.4 Hippocampal novelty-dependent responses

*"Selection is the very keel on which our mental ship is built. And in the case of memory its utility is obvious. If we remembered everything, we should on most occasions be as ill off as if we remembered nothing"*

(William James, 1890)

Distinguishing between what is novel and what is familiar, and distinguishing between degrees and types of novelty, is an elemental requirement for adaptive behaviour. As expected, there is massive behavioural evidence that animals routinely respond differentially to novel versus familiar stimuli. Electrophysiological recordings in humans and animals, as well as human neuroimaging studies, are beginning to elucidate the brain regions responsible for novelty detection. For

example, Tulving and colleagues (Tulving *et al*., 1996), using positron emission tomography (PET), demonstrated the existence of widely distributed neuroanatomical novelty-detection networks. Cortical and subcortical regions in the limbic system were more active in response to novel stimuli than otherwise comparable, familiar stimuli. As will be discussed later in this thesis, functional imaging studies of novelty responses have frequently demonstrated hippocampal responses to novel or unexpected stimuli.

Given James' reasoning that not all sensory input should be committed to memory, it follows that novel information should have preferential access to the memory systems. Therefore, it has been suggested that a function of novelty-sensitive brain regions is to determine the necessity of encoding information into long-term storage (Metcalf, 1993; Fabiani and Donchin, 1995; Tulving and Kroll, 1995). The novelty/encoding hypothesis developed by Tulving and Kroll (1995) conjectures that the encoding of incoming information depends on the novelty of that information. Put more strongly, novelty is a necessary, although not sufficient, condition for long-term storage of information. Furthermore, it is also suggested that novelty-assessment circuits serve the function of identifying on-line information whose encoding is of adaptive significance, and transmit it to other regions for further processing (Tulving and Kroll, 1995).

### *1.41 Behavioural responses to novelty*

Stimulus novelty is thought to elicit two states within an animal, curiosity or fear, with the ensuing behaviour considered a result of competition between these

two states (Montgomery, 1955). The dominant behavioural response to curiosity is exploration, while the response to fear is either withdrawal or immobility.

Intense stimulation, novel or not, tends to elicit a startle reaction, characterised by the adoption of a defensive posture (Fleschler, 1965). Less intense stimulation, which fails to elicit the defensive pattern of the startle reaction, may nevertheless elicit similar behaviour termed the arrest reaction, characterised by brief cessation of ongoing activities and autonomic change. The arrest reaction is rapidly followed by a complex of reactions labelled the orienting response. This response includes changes in muscle tone, respiration and circulation, neuroendocrine responses, desynchronisation of the cortical EEG, and orienting of the body and sense organs towards the source of stimulation (Sokolov, 1963). These reactions are thought to prepare the animal to register and analyse the source of stimulation. Following the orienting response, normal animals will either engage in overt exploration, moving towards and actively investigating the source of stimulation, or engage in some form of defensive reaction (Montgomery, 1955).

Few items or places are entirely novel. Novelty typically arises from new configurations of familiar elements. For example, in chapters 3 and 4, novelty-dependent responses were measured in response to letter strings. The letters themselves are, of course, highly familiar, yet they are presented in novel configurations. Novelty ‘wears off’ rapidly after several exposures; repeated exposures to items or places, with the opportunity to explore (O’Keefe and Nadel, 1978), leads to a growing sense of familiarity. This is reflected in a steady decrement in the behavioural response to a novel stimulus; a process known as habituation

(Groves and Thompson, 1970). In chapters 3 and 4, it is precisely this pattern of habituation, or, at a neuronal level, adaptation, with familiarity that will be measured. Hence, novelty seems dependent on long-term memory sensitive to contextual configurations and on the capacity to remember single occurrences.

That novel configurations of familiar stimuli are as effective at eliciting a novelty response as completely novel situations is illustrated by the findings of Wilz and Bolton (1971). Gerbils were allowed to explore, in an open field, either a group of objects in a particular spatial arrangement or a single object located in a specific place. The rearrangement of these objects, or positioning the single object in a new location, was as effective at eliciting exploration as a totally novel environment (Wilz and Bolton, 1971). Animal exploration of a novel environment usually involves gross bodily movements but can also be more subtle, mediated through visual search of the environment (Brown, 1968).

The orienting response (Sokolov, 1963) is not limited to novel stimuli; it is also evoked by salient stimuli. Salience refers to a property of those items or places that attract attention. Novelty almost always implies salience, although the converse is not the case. Examples of stimuli that attract attention are loud bangs, shiny objects, and blood, all of which are highly noticeable, although not necessarily novel (James, 1890). Salience can be considered as the extent to which the nervous system is excited, either because of stimulus intensity (e.g. loud bang) or its biological prepotency (e.g. blood). The critical variable controlling reactions to salience are recency and frequency of the same or similar occurrences. Hence, as with novel

stimuli, repeated exposure to salient stimuli leads to habituation, or adaptation, of the orienting response.

An item can also be salient because it violates the prevailing context in which it is encountered. The responses to such violations of context have been widely studied in 'oddball' paradigms (Rugg, 1995). Human neuronal responses to oddball stimuli are discussed in chapter 5. Oddball and other salient stimuli, as well as novel stimuli, elicit surprise. They are unpredictable, hence violate our predictive set about the world. In chapter 5, habituation of neuronal responses to repeated presentation of oddballs is discussed in terms of a repetition-dependent decrease in the unpredictability of an oddball stimulus.

*1.42 Effects of hippocampal lesions on behavioural responses to novelty*

Sokolov (1963) makes the important distinction between the orienting response and ensuing exploration. According to Sokolov, the neocortex maintains a model of stimuli in an environment. Novel stimuli produce a mismatch in this signal, which engages the reticular formation thereby producing an orienting response. As a model of the stimulus is built up in the neocortex, the mismatch signal decreases, leading to habituation of this orienting reflex.

The role of the hippocampus in the orienting response has been most extensively studied in rodents and remains a subject of controversy. O'Keefe and Nadel (1978) concluded that hippocampal-lesioned animals, even while engaged in some directed activity, show normal arrest and orienting responses to novel stimuli. However, these animals failed to explore the stimulus, returning instead to their

motivated task. Gustafson (1975) found that distraction durations to novel stimuli were lower in hippocampal rats and that this was due to reduced exploration coupled with normal orienting. O'Keefe and Nadel (1978) argue that exploration is driven by the occurrence of unpredicted events and that it ceases when the source of novelty is incorporated within the hippocampal representation of the environment.

Other authors have suggested, however, that hippocampal lesions impair habituation of the orienting reaction to novelty (Vinogradova, 1975). For example, Kim *et al*. (1970) found that the orienting response in hippocampal-lesioned rats either did not habituate or declined with great difficulty. Furthermore, hippocampal lesions impair habituation of the startle response to an acoustic stimulus (Mickley and Ferguson, 1989). Although Sokolov originally attributed mismatch detection to a neocortical-reticular formation interaction (Sokolov, 1963), he has also proposed that with repeated presentation of a novel stimulus, a neuronal model is created within the hippocampus with the time course of habituation reflecting the development of this neuronal model (Sokolov, 1990). This latter proposal unambiguously implicates the hippocampus in novelty processing.

Claims of impaired habituation were, however, disputed by O'Keefe and Nadel (1978) who pointed out that the hyperactivity of hippocampal-lesioned rats in novel environments is due to their lack of fear, their tendency to engage in repetitive, stereotyped behaviours, and their hyper-reactivity to stimuli. According to O'Keefe and Nadel (1978), activity patterns of hippocampal-lesioned rats did not reflect exploratory behaviour. Kimble (1975) has argued that increased activity following hippocampal lesions cannot simply be explained in terms of general hyperactivity

because it is most commonly observed in situations which are novel to the animal (open field, strange apparatus) or in which some significant element has been recently changed. Whether or not hippocampal lesions affect habituation of the orienting response, one conclusion that can be reached from both arguments is that these lesions do not inhibit the initial orienting response.

*1.43 Novelty-sensitive activity of single neurones*

*1.43a Rodents*

Early evidence for novelty processing in the hippocampus was provided by single-unit, extracellular, intracranial recordings in rabbits during stimulation with sensory stimuli from different modalities (Vinogradova, 1975). The majority of reactive neurons in CA3 responded to all stimulus types, indicating multimodality of input. With repeated presentations of a particular stimulus, the response duration gradually declined, and finally, by the $8^{th}$ to $20^{th}$ stimulus presentation, disappeared. This adaptive profile was not evident in entorhinal cortex and the response in dentate gyrus was non-decremental for a long period of stimulation, suggesting that the rapid habituation observed in CA3 was not secondary to a decrease in cortical input.

Any perceptible change of a signal, or the conditions of its presentation, caused the initial response to be reinstated. This could be evoked by a pause in stimulus application, change in its intensity, prolongation or shortening of the initial signal, change in the interval between rhythmically applied stimuli or of the number of stimuli given in a short series, and change in stimulus complexity (e.g. subtraction of a component from a standard complex). This led to the conclusion that CA3 neurons are sensitive to the novelty of a stimulus (its absence in “trace storage”), not

its specific physical properties. Furthermore, Vinogradova claimed that neuronal responses of habituation and dishabituation of CA3 responses were similar to the dynamics of the orienting response at the behavioural level.

Responses in hippocampal CA1 neurones were found to be much more stimulus-specific than those evoked by CA3 neurones (Vinogradova, 1975). The peculiar feature of CA1 neuronal dynamics was the absence of the reaction to the first stimulus in a series, with the maximal response being evoked by the second or third stimulus. Adaptation occurred rapidly after the stimulus was presented about 8 times. These cellular responses led Vinogradova to speculate that CA1 neurons, in addition to responding on the basis of stimulus novelty, are, because of the stimulus-specificity of their responses, able to code the quality of incoming sensory information. The source of this quality was suggested to be a direct cortical-CA1 projection. It was also suggested that the CA3 acts as a trigger or gate, which only in some definite state (novelty, mismatch) allows passage of sensory signals through to CA1 neurones. A caveat in the interpretation of Vinogradova's results is that, unfortunately, no record of the animal's behaviour or EEG was taken, and others have claimed that these results could be due to arousal from sleep (Mays and Best, 1975).

O'Keefe and Nadel (1978) characterised the units in the CA1 field of the hippocampus of the freely moving rat into two general classes: place units and displace units. Displace units are those whose firing pattern relates to the behaviour of the animal, irrespective of where it occurs in an environment. The firing pattern of place units is dependent on the exact location of the animal in an environment. The

part of the environment where the unit fires, or fires maximally, is called the place field. For some place units, the rat's location in an environment is a necessary, but not sufficient, condition for unit firing; in addition, the animal must be behaving in a specific way or receiving a specific stimulus. Included in this last category of complex place units are misplace units which fire maximally when a rat attends to a place either because the object usually there (e.g. reward) has been removed or, alternatively, because a novel object has been placed there. Ranck (1973) also described hippocampal mismatch cells. These approach-consummate-mismatch units, found primarily in CA1, fired maximally when a rat went to a reward site and failed to find reward. Hence, in contradistinction to mismatch sensitivity found by Vinogradova (1975) in CA3, the models developed by O'Keefe and Nadel (1978) and Ranck (1973) localised mismatch detection to the CA1 field.

Gray (1982) also developed a theory of hippocampal mismatch detection. His comparator theory of septo-hippocampal function, derived primarily from behavioural experiments, suggests a role for the subiculum as a central comparator. Anatomical studies demonstrate that sensory information is sent simultaneously from the entorhinal cortex to the dentate gyrus and subicular area (see chapter 7). Vinogradova's (1975) findings show that the single-cell responses to simple sensory stimuli are less specific to the parameters of the stimulus in the dentate gyrus than in entorhinal cortex. Gray's theory addresses the functional significance of this dual projection; specifically it addresses the question of why entorhinal cortex sends less specific information by a more circuitous route, via the dentate gyrus and CA subfields, to the subicular cortex, to which it also apparently sends more specific information directly. Gray suggested that the dentate cells and CA subfields serve as

a match-mismatch system to determine whether the information is ‘important’ to the animal and, in the case of ‘mismatch’, send an enabling signal from CA1 to the subiculum cells. Hence, like Vinogradova (1975) and O’Keefe and Nadel (1978), mismatch detection occurs in the CA subfields. In Gray’s theory, the direct sensory information to the subiculum is used for matching against predicted sensory input, and for generating the next prediction, but only if it receives an enabling signal from CA1. Thus "direct input to the subicular area from the entorhinal cortex describes the current state of the world, while the input via the hippocampus determines whether the description is treated as important" (Gray, 1982; p. 271).

In summary, the results from single unit recordings in rodents suggest that components of the hippocampus are sensitive to mismatches between expectation and experience. It is not clear, however, whether the hippocampus is critical for evoking the orienting response towards novel or salient stimuli. Importantly, O’Keefe and Nadel (1978) and Gray (1982) both state that mismatch detection in the hippocampus serves to initiate specific explorative and investigative behaviour.

*1.43b Rodent EEG oscillatory activity*

Theta activity is a widely studied component of the hippocampal electroencephalogram (EEG). Theta is a slow sinusoidal rhythm ranging from 6-10 Hz in the rat. The circuitry of the hippocampus is tuned to oscillate at theta frequencies and these oscillations are driven by pulsed inputs from the medial septum and diagonal band of Broca. In addition, it is possible that the projection from the hippocampus to the septum is itself important for triggering hippocampal theta activity. The hippocampal EEG can be related not only to neuronal mechanisms

but also to the behaviour of the animal. A large number of experiments have been carried out to determine the behavioural and psychological correlates of these EEG states, particularly theta activity.

Theta dominates the hippocampal EEG during behaviours elicited by novelty or discrepancy: exploration of a novel environment, the orienting reflex elicited by a novel stimulus and general searching behaviour. Theta is elicited when an animal is first placed in a novel, unfamiliar environment, when a new stimulus is placed in a familiar environment, when some aspect of a familiar environment is changed, for example when the reward is omitted during the extinction phase of a learning task, and when an animal appears to be searching for something. Hence theta is evoked under similar circumstances to those causing mismatch cells to fire (Ranck, 1973; Vinogradova, 1975; O'Keefe and Nadel, 1978). Theta has, however, also been associated with motivation, emotion and voluntary movement (Bennett, 1975), making it difficult to ascribe theta one precise functional role.

Theta oscillatory activity pertains to rhythmical changes in electrical potential within the neuronal cell body. These oscillating potential changes are termed sub-threshold because they are below the threshold potential change required to evoke an action potential. The relationship between oscillatory activity and the signal measured in fMRI (see chapter 2 part I) is at present unknown. However, being in a state of theta may itself facilitate stimulus-induced changes in neuronal firing that are detectable by fMRI. For example, Buzsaki *et al*. (1981) described how information about event unexpectedness might be modulated by hippocampal theta. They found

that perforant pathway stimulation evokes the greatest responses from the dentate during the negative going phase of hippocampal theta.

*1.43c Monkey*

Electrophysiological recordings from medial temporal lobe of monkeys performing recognition memory tasks using large stimulus sets have repeatedly demonstrated neurones that respond less during subsequent presentations of visual stimuli previously encountered (see Brown and Xiang, 1998 for review). The reductions in neuronal responses signal either the relative familiarity or relative recency of that stimulus. Response reductions occur commonly (~25% of recorded neurones) in anterior temporal cortex, notably the perirhinal cortex, but are much less frequent (<~1%) in the hippocampus (Brown *et al*., 1987; Riches *et al*., 1991; Miller *et al*., 1993; Rolls *et al*., 1993; Xiang and Brown, 1998). It should be noted, however, that the proportion of neurones in the hippocampus that are simply visually responsive is considerably less than in other recorded regions (Xiang and Brown, 1998), perhaps suggesting that visual input, and, in turn, relative familiarity of visual input, is sparsely coded within the hippocampus.

Response suppression in perirhinal cells shows single-trial learning for individual visual stimuli and long-term maintenance (>24 hrs) of this differential response (Brown and Aggleton, 2001). By contrast, the few hippocampal neurones that have been found to change their response after repetition of individual stimuli show response changes of much longer latencies than those of perirhinal neurones and these differential responses do not persist for 24 hrs (Rolls *et al*., 1993; Xiang and Brown, 1998). The majority of novelty-sensitive neurones recorded in the

macaque hippocampus by Rolls *et al*., (1993) had memory spans of finite length (median 21 intervening stimuli), which led to the conclusion that their activity related to recent memory rather than to memory of whether a stimulus had ever been seen before. Vinogradova (1975) demonstrated response changes on repetition of simple stimuli in rabbit hippocampus (section 1.43a) but did not investigate the possible memory span of these responses.

There is, therefore, strong evidence from monkey recordings that the crucial processing in judging prior occurrence occurs in perirhinal cortex (and adjacent visual association cortex; area TE), independent of hippocampus. The latency for discrimination of prior occurrence is as fast as the latency for identification of the visual stimulus within monkey perirhinal cortex and anterior area TE (Miller *et al*., 1993; Xiang and Brown, 1998). The speed of this response makes it unlikely that it is mediated by top-down influence from hippocampus. The long-latency and sparsity of hippocampal novelty-dependent responses (Rolls *et al*., 1993; Xiang and Brown, 1998) accords with this lack of hippocampal dependency.

Medial temporal cortical responses are typically stimulus-specific, responding differentially to relative familiarity of certain stimuli and not others (Young *et al*., 1997). Novelty responses in hippocampal cells, however, do not show this stimulus-selectivity. Novelty-sensitive neurones typically respond to all novel stimuli (Vinogradova, 1975; Rolls *et al*., 1993; Otto and Eichenbaum, 1992; Wiebe and Staubli, 1999) suggesting that these neurones don't themselves represent stored information. Rolls *et al*., (1993) suggested that these neurones reflected the read out from a memory store. Alternatively, it could be suggested that these neurones reflect

processing of generic novelty i.e. they are sensitive to any unexpected stimulus not previously encountered, perhaps suggesting that the hippocampus mediates abstracted, stimulus-general mismatch detection.

Neurones within the monkey and rodent hippocampus have been shown to encode information about the relative familiarity of a visual stimulus occurring in a particular spatial location (Rolls *et al*., 1989; Wood *et al*., 1999; Eichenbaum, 2000). By contrast, perirhinal neurones are much less involved in encoding allocentric space (Burwell *et al*., 1998). Furthermore, there is evidence from immediate early gene imaging in rats that demonstrate hippocampal, but not perirhinal, engagement in response to novel spatial arrays of stimuli (Wan *et al*., 1999). Hippocampal neurones may, therefore, provide a potential substrate for recognition memory processes involving spatial and other associative information.

*1.43d Human*

There have been few experiments examining single unit responses to novelty in the human hippocampus. Recordings are typically obtained from electrodes placed in the medial temporal lobes of epileptic patients. The single unit recordings, as well as local field potentials that can be recorded from these electrodes, provide important evidence for the temporal profiles of local neuronal processes in human hippocampus. However, generalisation about normal neuronal function from recordings in epileptic patients constitutes a potential limitation. Fried *et al*. (1997) demonstrated that single neurones in the human hippocampus fire differentially to novel versus familiar stimuli. These differential responses were observed as long as

10 hr after initial stimulus presentation. It is not clear, however, whether differential firing of these neurones reflected recognition memory processes (Brown and Aggleton, 2001) as they were not only sensitive to relative familiarity, but responded to conjunctions of stimulus attributes and stimulus novelty or familiarity. Heit *et al*., (1988) failed to find novelty sensitive units in the human hippocampus but Rolls *et al*., (1993) suggested that the negative result obtained by Heit and colleagues could reflect upon the relatively small numbers of novelty-sensitive neurones present in the primate hippocampus.

*1.43e Human local field potentials*

Stronger evidence for a human hippocampal role in novelty processing comes from event-related field potential recordings within the medial temporal lobes in epileptic patients. Field potentials reflect the summation of local neuronal population firing. In an experiment examining the effects of repetition of words on medial temporal potentials, Grunwald *et al*. (1998) demonstrated that damage to the hippocampus proper (patients with sclerosis) attenuated anterior medial temporal lobe event-related potentials for novel visually presented words. Responses to repetitive presentations were unaffected. The authors concluded that the hippocampus is sensitive to the first appearance of a known verbal stimulus in a new situation, i.e. situational novelty of verbal stimuli.

A previous study had demonstrated hippocampal sensitivity to infrequent, unpredicted stimuli. Halgren *et al*. (1980) demonstrated large, long-latency field potentials elicited in the human hippocampus by rare tones in an auditory oddball task. These potentials were labeled medial temporal P3 potentials because they were

elicited by the same task conditions that elicit the P3 potential recorded at the scalp (the P3 event-related potential is discussed in chapter 5). Similar potentials were evoked by rare visual stimuli, randomly intermixed with frequent visual stimuli of equal luminance, colour and contrast. These potentials could also occur in the absence of an evoking sensory stimulus, as when a tone is occasionally omitted from a regular series of identical tones. Hence, like the novelty responses recorded by Vinogradova (1975) in the rabbit, these responses did not depend on sensory modality or quality of the stimulus. The responses were enhanced when the stimulus was rare and attended. Evidence that these evoked potentials correlate with firing of hippocampal neurones comes from studies that simultaneously measure single-unit action potentials and local evoked field potentials (Squires *et al*., 1976; Halgren *et al*., 1983).

Donchin (1981) suggested that larger P3s are elicited by infrequent tones in the auditory oddball task because they provoke a larger shift in the mental context that has been ingrained by repetition of the more frequent tones. The importance of the *shift*, rather than the identification of rarity *per se*, is supported by the results of Squires *et al*. (1976; see also Halgren *et al*., 1983), where medial temporal P3s are elicited by either high or low tones when they signal a shift from the preceding sequence, even though the overall frequency of occurrence of high and low tones is equal. Although the authors do not comment upon this, a similar finding can be observed in single-unit human hippocampal intracranial recordings (figure 1 in Kreiman *et al*., 2000). Repeating the same stimulus (A) causes response adaptation. Following the presentation of a few intervening stimuli (B), the response to A recovers to slightly below its original magnitude. A similar effect of intervening

stimuli, described above, was observed by Rolls *et al*., (1993) in the macaque hippocampus. Specifically, the magnitude of hippocampal neuronal response recovery was found to correlate with the number of intervening stimuli i.e. as the number of intervening stimuli between first and second presentation of a stimulus increased, the neuronal response increased to a level similar to that observed for novel stimuli.

One conclusion from single unit recordings, in all 3 species, is that hippocampal responses index mismatch between expectation and experience. A further role for the hippocampus might reflect recognition processes for spatial arrays of stimuli (Brown and Aggleton, 2001), although this has yet to be directly tested in humans. Properties of perirhinal responses enable judgements of prior occurrence of simple stimuli on the basis of familiarity.

*1.44 How other models of hippocampal function may suggest novelty-dependent processes in the hippocampus*

In the 1980s, two influential theories of hippocampal function, based on animal models, were the spatial mapping theory of O'Keefe and Nadel (1978) and the working memory theory of Olton and colleagues (Olton *et al*., 1979). Olton's working memory (WM) theory proposed that the hippocampus is selectively concerned with maintaining information that is pertinent only within a short period of time (working memory; Baddley, 1992) rather than information that remains constant over time (reference memory; RM). Thus, the hippocampus is seen as being concerned not with the spatial or nonspatial nature of information but rather with how the information is processed in memory. However, it has been pointed out that

the distinction between WM and RM is often not clear-cut because WM procedures always incorporate some element of RM (Morris, 1983). It could even be speculated that the WM/RM distinction actually refers to the episodic/semantic distinction described earlier in this chapter. Of interest, however, is that Olton's working memory hypothesis suggested a hippocampal role in processing dynamically changing, recent events. This could be reinterpreted as hippocampal sensitivity to change or a hippocampal role in directing switches in behaviour.

Olton's task (radial maze) bears resemblance to spontaneous alternation, which refers to the following phenomenon. If a sated rat is given two consecutive trials in a two-choice apparatus, such as a T-maze, it is highly likely that its second choice will differ from the first. Rats with hippocampal lesions fail to alternate choices in spontaneous alternation conditions (Gaffan, 1972). Spontaneous alternation demonstrates that some trace of the first choice is subsequently available to the rat and furthermore, the rat is in some way motivated to change its behaviour on the second trial. The fact that hippocampal lesions impair alternation lends support to a hippocampal role in directing switches in behaviour, but could equally be explained by an inability to remember the previous choice.

In the spatial mapping theory, the hippocampus is seen as being specifically involved in spatial learning and memory. Some of the most convincing data supporting this theory comes from hippocampal single-unit recordings made during exploration (O'Keefe and Dostrovsky, 1971; Ranck, 1973, O'Keefe, 1976). Specifically, recordings made from CA1 and CA3 cells show that these cells fire preferentially when the rat is in specific places in the environment. Furthermore,

extensive removal of rat hippocampus produces impairments on place learning tasks (e.g. Morris *et al*., 1982). In humans, functional imaging studies of spatial navigation have consistently demonstrated hippocampal activation (see chapter 7).

Critical evidence for the spatial mapping theory comes from the deficits displayed by hippocampal-lesioned rats in the Morris water maze task. In this task, rats or mice learn to escape from submersion in a pool by swimming to a hidden platform located just underneath the water surface. Importantly, training in the conventional version of the task involves an intermixing of four kinds of trial episodes that differ in the starting point of the swim. Under this condition, animals with hippocampal damage typically fail to acquire the task (Morris *et al*., 1982). However, if the demand for synthesising a solution from four types of episodes is eliminated by allowing the animal to repeatedly start from the same start position, animals with hippocampal damage acquire the task almost as readily as normal rats and use the same distant spatial cues to identify the escape site (Eichenbaum *et al*., 1990). However, when the rats with hippocampal damage that have successfully learned to locate the escape platform from a single start position are tested from new positions, they fail to readily locate the platform. In contrast, normal animals that were previously trained from one start position swim directly to the escape locus from each new starting location (Eichenbaum *et al*., 1990). Whishaw and Tomie (1997) obtained the same result and, furthermore, demonstrated that the impairment following hippocampal lesions was aggravated by perseverative returns to the first learned place.

It is thought that hippocampal-lesioned rats fail to develop an allocentric representation of the water maze that is independent of start location. Alternatively, it could be argued that hippocampal lesioned rats fail to detect a change in platform location or they may be impaired in switching their responses, i.e. hippocampal rats may perseverate despite normal novelty detection. Hippocampal lesion-induced impairment on new learning has been demonstrated in many kinds of tasks in both primates (Angeli *et al*., 1993) and rats (see O'Keefe and Nadel, 1978 for review) and has been referred to as a reversal deficit.

In the human literature, bilateral hippocampal damage does not produce perseveration on the Wisconsin Card Sorting Task (WCST). The WCST is a series of visual discriminations across multidimensional stimuli, in which the rule governing reinforcement is periodically changed across different dimensions of the stimuli (Grant and Berg, 1948; see chapter 4). HM was able to detect abstract, task-related rule changes and modify his responses accordingly (Milner, 1963). The WCST requires subjects to remember the current rule, which is thought to engage working memory (Baddley, 1992). The WCST does not, therefore, involve response switching in the context of episodic memory encoding.

Perseveration of responses does arise when human amnesics engage in episodic memory tasks that provoke proactive interference. Warrington and Weiskrantz (1970, 1974) demonstrated very poor free recall and recognition for words in amnesics, but relatively normal cued recall. Providing amnesics with a degraded version of the target word, or its initial letters, improved patient performance differentially with respect to controls. Similarly, Winocur and

Weiskrantz (1976) showed good amnesic learning of paired associates that were related by either a semantic or rhyme rule. Although amnesics showed good cued recall under the above circumstances, they were impaired on cued recall of a second list presented a few minutes after the first (Warrington and Weiskrantz, 1974, 1978; Winocur and Weiskrantz, 1976). The second list in these reversal learning studies consisted of either the same initial three letters as those in the first list (Warrington and Weiskrantz, 1974, 1978) or of word pairs that were linked by the same rule but in which the response differed from that of the first list (Winocur and Weiskrantz, 1976). Like hippocampal lesioned rats returning to the first learned platform location (Whishaw and Tomie, 1997), human amnesics in the above two reversal learning tasks made many intrusion errors. Warrington and Weiskrantz (1978) interpreted their results in terms of slower unlearning of previous responses in amnesics. These findings, together with HM's normal performance on the WCST, suggests that detecting changes in task demands and directing changes in behaviour is intact following medial temporal damage provided the task does not require updating the contents of episodic memory. These findings are discussed further in chapter 7.

The hippocampal lesion-induced impairment in new spatial learning in the Morris water maze described by Eichenbaum *et al*. (1990) could also be interpreted as a deficit in applying previously learned information to a novel situation. Other lines of experimental evidence suggest that the hippocampus is required for utilising previous knowledge for new problem solving. In several experimental protocols designed by Eichenbaum and colleagues, animals with hippocampal damage successfully acquire a set of overlapping experiences, often at a rate not significantly different from that of normal subjects. The lesioned animals, however, fail to express

their memories of the experience in new situations that require an inference on the basis of linking the distinct experiences in memory. Eichenbaum (2000), like Pribram and Isaacson (1975), concluded that in animals, the role of the hippocampus is the flexible expression of memories.

An early example of flexible expression of hippocampal-dependent memory comes from Eichenbaum *et al*. (1989). Rats with hippocampal damage were trained concurrently on two separate odour discrimination tasks (A+B- and C+D-) that they could eventually perform about as well as normal rats. Thus, both normal rats and rats with lesions came to chose odour A when it was presented in the odour pair AB and odour C when it was presented in odour pair CD. However, a transfer task showed that something different had been learnt by the two groups. Specifically, when rats were presented with recombined odour pairs (AD or BC), the normal rats tended to chose odour A, performing about as well as on the regular learning trials. They were not disrupted by the new combination of stimuli and were able to use relational information about the odours in a flexible way. In contrast, the rats with hippocampal lesions behaved as if they were confronted with a new problem and performed near chance. In their case, it appeared that the kind of knowledge that had been acquired was inaccessible when the original learning event was not precisely repeated.

In another study, rats learned overlapping sets of associations between odour stimuli (Bunsey and Eichenbaum, 1996). On each trial, one of two odours was initially presented, followed by a choice between two odours, one of which was baited as the assigned ‘associate’ for a particular initial odour (A goes with B, not Y;

X goes with Y, not B). Following training on two sets of overlapping odour-odour associations (A-B and X-Y, then B-C and Y-Z), subsequent probe tests were used to characterise the extent to which learned associations could be linked to support inferential memory expression. Control rats learned paired associates rapidly and hippocampal damage did not affect acquisition rate on either of the two training sets. Intact rats also showed that they could link the information from overlapping experiences, and use this information to make inferential judgements in two ways. Firstly, normal rats showed strong transitivity across odour pairings that contained a shared item. For example, having learned that odour A goes with odour B, and that B goes with C, they could infer that A goes with C. Secondly, control rats could infer symmetry in paired associate learning. For example, having learned that B goes with C, they could infer that C goes with B. By contrast, rats with hippocampal lesions were severely impaired, showing no evidence of transitivity or symmetry (Bunsey and Eichenbaum, 1996).

These studies demonstrated that some forms of stimulus-stimulus representations can be acquired independently of the hippocampus. However, these representations are 'hyperspecific' in that they can only be expressed within the confined context of the reproduction of each of a set of distinct learning events (see earlier discussion of declarative versus nondeclarative memory). On the basis of these animal studies, Eichenbaum (2000) has suggested that only hippocampally-mediated representations can support the inferential expression of associations that must be linked across separated experiences. However, because hippocampal rats are impaired if they have to apply previously learned information to a novel situation, an

alternative suggestion could be that hippocampal rats do not correctly acknowledge or process the novelty of these new learning situations.

In humans there is evidence that the novelty processing function of hippocampus may be even more fundamental than its role in explicit vs implicit memory (see above). A recent meta-analyis has shown that amnesics perform worse than controls in tests of implicit memory (thought to be independent of hippocampus) if the study material is novel (Gooding et al., 2000). Implicit memory for familiar information was preserved in the amnesics.

## 1.5 Functional imaging as a tool to investigate memory processes of the human hippocampus

Early attempts to detect hippocampal activation using PET or fMRI largely failed (reviewed by Fletcher *et al*., 1997). This difficulty in finding hippocampal activation was commented upon in early papers, with authors proposing various technical, neurobiological and psychological explanations. The medial temporal lobe is subject to fMRI susceptibility artifacts and signal drop-out (Ojemann *et al*., 1997), yielding a smaller signal to noise in the medial temporal region compared to other cortical regions. The possibility was also raised that memory encoding or representation by the hippocampal system is so sparse (i.e. circumscribed in the portion of the hippocampal network engaged by any learning event) that it would produce very small activations. A different set of concerns related to the possibility that functional imaging required different paradigms to those used in

neuropsychology and, perhaps more drastically, a reconceptualisation of the nature of the memory processes mediated by the MTL.

The number of reports of hippocampal activation in various memory tasks that have been published over the last five years (see Lepage *et al.*, 1998; Schacter and Wagner, 1999 for review) dispels the first two concerns, enabling functional imaging to be used as an additional tool for exploring the brain bases and functional organisation of memory. Furthermore, the findings from functional imaging have found significant concordance with neuropsychological data. This concordance is particularly apparent in terms of the range of to-be-remembered material and hippocampal material-specific laterality.

Studies of HM and other patients with bilateral hippocampal lesions show that amnesia is a global memory deficit. Such patients have material- and modality-general impairments, encompassing verbal and nonverbal, spatial and nonspatial stimuli, regardless of whether they are presented visually or auditorily (Milner and Teuber, 1968), indicating that the hippocampal role in memory is non-specific with regard to material and modality. However, studies of patients with unilateral damage to the left or right medial temporal lobe show clear material-specific memory deficits: verbal and nonverbal memory performances are selectively compromised following medial temporal damage in left and right hemispheres, respectively (e.g. Milner, 1972). Thus there is laterality to the hippocampal contribution to memory, corresponding to the types of processing for which the hemispheres are specialised.

Results from functional imaging studies demonstrate both the globalness of hippocampal processing, when considered bilaterally, and material-specific laterality. Two recent studies have addressed laterality directly. Kelley *et al.* (1998) presented subjects with words, nameable line drawings of objects, and unfamiliar faces in separate blocks, with instructions to either passively view or study the items for a later memory test. Independent of task demands, significant hippocampal activation, relative to viewing a fixation point, occurred in all three conditions, with hemisphere engagement varying as a function of stimulus type: words engaged left hippocampus, unfamiliar faces predominantly right hippocampus and nameable line drawings resulted in bilateral hippocampal activation. A similar PET study (Martin *et al*., 1997) demonstrated that, compared to viewing visual noise patterns, viewing of objects produced more right hippocampal activation than left, whereas viewing words produced more left hippocampal activity than right.

The range of studies activating the hippocampus indicates the broad range of stimulus types that engage the hippocampus. Regardless of which hemisphere is preferentially activated, hippocampal activation has been reported for words, visually (Martin *et al*., 1997; Kelley *et al*., 1998; Brewer *et al*., 1998; Wagner *et al*., 1998) or auditorily presented (Dolan and Fletcher, 1997; Kopelman *et al*., 1998), objects (Schacter *et al*., 1995; Kanwisher *et al*., 1996; Martin *et al*., 1997; Kelley *et al*., 1998), scenes (Tulving *et al*., 1996; Stern *et al*., 1996; Gabrieli *et al*., 1997; Brewer *et al*., 1998; Montaldi *et al*., 1998), faces (Sergent *et al*., 1992; Grady *et al*., 1995; Kapur *et al*., 1995; Haxby *et al*., 1996; Kelley *et al*., 1998), spatial routes (Maguire *et al*., 1997), and landmarks or locations (Maguire *et al*., 1997; Aguirre and D'Esposito, 1997).

The above studies illustrate the effectiveness of functional imaging as a tool for investigating hippocampal function. Further details of these studies will be presented in later chapters. One criticism, however, of functional imaging studies of the medial temporal region is that authors have used the term hippocampus, or hippocampal region, to loosely describe activations anywhere in the medial temporal lobe. As described in this chapter, different components of the medial temporal region have been shown to possess different functional properties (e.g. hippocampus vs perirhinal cortex). Even within the hippocampus, there is now a suggestion that hippocampal subregions in animals (Moser *et al*., 1993) and humans (Lepage *et al*., 1998; Schacter and Wagner, 1999) mediate distinct aspects of memory. Evidence for functional segregation within the medial temporal lobes emphasise the importance of detailed anatomical specification in functional imaging studies of the hippocampus.

### 1.6 Overview of thesis

The emphasis in this thesis is the study of novelty processing in the human hippocampus as revealed by functional neuroimaging. The outlined approach is predicated on an assumption that objects or items in the world are rarely completely novel. Indeed the general assumption is that there are two principal components to novelty. The first of these relates to physical qualities of stimuli and recognises that, in most instances, novelty arises from new configurations of familiar elements. Secondly, novelty also has a temporal component and within this perspective it is defined as recency of prior occurrence of a stimulus, a definition which embodies the idea that novelty responses have a time-constant. The overall biological question

addressed is whether the hippocampus is sensitive to novelty and is addressed in a series of experiments that first ask whether the hippocampus is sensitive to recency of prior occurrence. I then expand on this question by characterising adaptation of hippocampal novelty responses with repeated stimulus presentations. A related issue that is addressed concerns whether a hippocampal response to novelty is dependent on the behavioural relevance of the stimulus. I then test the hypothesis that the critical variable for evoking novelty-dependent activity in the hippocampus is mismatch between expectation and experience. Finally, the role of the hippocampus in episodic memory encoding is addressed and the possible relationship between novelty detection and encoding is discussed.

In this chapter the multiple memory systems of the brain have been introduced. In particular, the role of the human hippocampus in episodic memory has been described, with emphasis on the hippocampal role in detecting novel vs familiar stimuli. The review of functional imaging studies of the hippocampus demonstrates the suitability of fMRI for measuring responses in this brain region and also illustrates the fact that novelty detection is one of several reported functions of the human hippocampus.

In chapter 2 the principles of fMRI are introduced. The physics of magnetic resonance are described as well as the neurophysiology underlying the blood oxygen level-dependent (BOLD) signal, which is the measure of neuronal activity in the experiments described in this thesis. The second part of chapter 2 outlines the analysis of fMRI data. Spatial preprocessing of fMRI images is described as well as

the statistical models employed in making inferences about task-related regional brain activations.

Chapters 3 to 6 describe a series of experiments investigating the types of novelty the human hippocampus is sensitive to, as well as the behavioural conditions under which novelty responses are elicited. In chapter 3, novelty is introduced in the context of an item-learning paradigm generated from an artificial grammar system, an arbitrary set of rules governing the concatenation of symbols (Reber, 1967). This enabled a test of whether hippocampal novelty responses depend upon behavioural relevance. The experiment presented in chapter 4 expands on the findings of chapter 3. In chapter 4 novelty is introduced in the context of explicit rule learning. By contrast to chapter 3, in which the rule system is constant throughout the experiment, the rule governing behaviour in chapter 4 is periodically changed. This enabled a test of whether the hippocampus is sensitive to a high level, abstract form of novelty.

Chapter 5 examines oddball-evoked neuronal activity, that is, responses to stimuli that violate their prevailing context. Specifically, responses to 3 types of oddballs, perceptual, semantic and emotional, were measured in order to investigate activity commonly evoked by different oddball types. The second part of chapter 5 investigates a hippocampal role in detecting mismatches between expectation and experience by testing for hippocampal responses to oddballs that adapt with repeated presentations of oddballs.

Chapter 6 investigates medial temporal contributions to successful episodic memory encoding with fMRI scanning parameters manipulated so as to maximise

sensitivity to responses in anterior medial temporal lobe. During verbal encoding, responses to words subsequently recalled are compared to responses to subsequently forgotten words.

The conclusion reached from chapters 3 to 6 is that the functional properties of anterior and posterior segments of hippocampus dissociate. Anterior hippocampus is functionally specialised for detecting mismatches between expectation and experience. By contrast, posterior hippocampal responses index stimulus familiarity, which may reflect retrieval from episodic memory. Chapter 7 proposes possible bases for this functional segregation. A review of the literature suggests that distinct connectivity profiles and segregated inputs from neuromodulatory systems may give rise to different functional properties of discrete hippocampal regions along the longitudinal axis. A general discussion of the experimental findings in this thesis is presented in chapter 8.