Gracias a la vida que me ha dado tanto
me dio dos luceros que cuando los abro
perfecto distingo lo negro del blanco
y en el alto cielo su fondo estrellado

Thanks to life which has given me so much
it has given me two eyes, and when I open
them I clearly distinguish black from white
and in the sky, its starry depths

Gracias a la vida que me ha dado tanto
me ha dado el oído que en todo su ancho
graba noche y día grillos y canarios
martirios, turbinas, ladridos, chubascos

Thanks to life, which has given me so much
it has given me hearing, which in all its breadth
day and night records crickets and canaries,
hammers, turbines, barking, dark clouds

Gracias a la vida que me ha dado tanto
me ha dado el sonido y el abecedario
con él, las palabras que pienso y declaro

Thanks to life, which has given me so much
it has given me sound and the alphabet
and with it the words to think and speak

Gracias a la vida que me ha dado tanto
me dio el corazón que agita su marco
cuando miro el fruto del cerebro humano

Thanks to life, which has given me so much
it gave me my heart, which shakes its frame
when I look at the fruit of the human brain

Gracias a la vida, gracias a la vida
Por Violeta Parra

Thanks to life, Thanks to life
By Violeta Parra
Acknowledgements

This thesis is dedicated to all those who think, to those who devote their energy to the cause of understanding the underpinnings of the human mind, to all those who inspire and motivate me to contribute to this inquisitive and fascinating enterprise. In particular, I would like to thank those who have provided the personal support during the production of this thesis, my family and friends. I thank my supervisor, Ray Dolan, for his constancy and patience. My warmest gratitude also to those who have shared their various kinds of knowledge with me: Christian Büchel, Stefan Kiebel, Jorge Armony, Rik Henson, Oliver Jospehs, Geraint Rees, John Ashburner, Andrew Holmes, Tim Shallice, and all the authors of neuroscientific and psychological literature from whose work I have benefited. I thank my subjects for their readiness to participate and for providing the data upon which this thesis is based.
The aim of this PhD is to describe the functional neuroanatomy underlying associative learning processes, specifically those involved in the crossmodal integration of information, using functional magnetic resonance imaging. The first study (chapter 3) investigated brain systems engaged during learning word paired- associates with mediation of mnemonic strategies, such as imagery (nonverbal) and semantic relatedness (verbal). This study revealed activation in visual cortex mediating use of nonverbal cues during encoding and in frontal and temporal cortices reflecting verbal and semantic manipulations of the material. The second and third experiments (chapters 4 and 5) studied the neural systems involved in making associations between abstract auditory (music chords) and visual (Chinese ideograms) stimuli. In these studies, hippocampus and insula-claustrum activations were observed during crossmodal associative encoding. The fourth study (chapter 6) investigated modulatory responses within sensory cortices as a function of associative learning. Here it was possible to show that through associative processes, auditory cortex can respond to visual stimuli (colours), which become predictive of auditory stimuli (meaningless sounds). The reverse was also observed in visual cortex. The fifth experiment (chapter 7) compared brain regions mediating crossmodal transfer (encoding and retrieval in different modalities) versus intramodal (encoding and retrieval in the same modality) retrieval of information in the visual and auditory modalities. The question addressed in this experiment was how information, which is represented in various modalities, is integrated and subsequently evoked by another input modality. The main finding in this case was the crucial involvement of left inferior frontal gyrus in crossmodal retrieval. This thesis provides a description of brain systems involved in associative learning, with specific stress on crossmodal associations. The findings suggest such brain systems involve primary and association sensory cortices as well as insula-claustrum, multisensory temporoparietal areas, and association prefrontal cortex.
CHAPTER 3: A NEUROANATOMY OF MNEMONICS ...........82

INTRODUCTION ................................................................. 82

METHODS ........................................................................... 86
Subjects ............................................................................. 86
Experimental materials ..................................................... 86
Cognitive task .................................................................... 87
Experimental Procedure ................................................... 87
Magnetic Resonance Procedures ........................................ 88
Data analysis ..................................................................... 89

RESULTS ........................................................................... 90
Behavioural data ............................................................. 90
Neuroimaging data .......................................................... 91

DISCUSSION ................................................................. 102
Conclusion ....................................................................... 109

CHAPTER 4: CROSSMODAL ASSOCIATIVE LEARNING I ....111

INTRODUCTION ................................................................. 111

METHODS ........................................................................... 113
Subjects ............................................................................. 113
Experimental materials ..................................................... 114
Cognitive task .................................................................... 114
Experimental Procedure ................................................... 115
Magnetic Resonance Procedures ........................................ 116
Data analysis ..................................................................... 117

RESULTS ........................................................................... 118
Behavioural data ............................................................. 118
Functional Neuroimaging data .......................................... 120
a) Main effects ................................................................ 120
b) Time-dependent interactions ........................................ 120
c) Modulatory response in extrastriate cortex .................... 127
CHAPTER 1: General Introduction

The work presented here deals with functional neuroanatomical descriptions of brain systems involved in associative learning, in particular the formation of associations between stimuli in different sensory modalities in humans. The interest in this topic arises from the ecological relevance of learning to integrate various sources of sensory information into related or unified representations. The specific aim of the thesis is to better understand how the human brain performs such perceptual and cognitive operations. The use of functional neuroimaging for this purpose allows the localisation and description of the time-course of brain activations, which participate in the relevant stages of information manipulation and integration. Processing stages, such as unimodal sensory processing, convergence of inputs from different sensory modalities, retrieval of previously acquired information, and the creation of new associations between stimuli or between stimulus and context, can be characterised with reference to brain activations by careful experimental specification of these processes.

The studies presented in this thesis are all based on the method of blood oxygenation level-dependent (BOLD) functional magnetic resonance imaging (fMRI), which enables the correlation of neural responses, in particular brain areas, with underlying cognitive or perceptual operations. FMRI also affords the possibility of assigning specific neural responses to single item presentations (in the sense that event-related fMRI allows observation of BOLD-signal in response to single events). This approach, known as event-related fMRI, is especially useful in studies where items have to be separated a posteriori according to certain criteria, such as subjects' performance. For instance, in a learning experiment, scans corresponding to presentation of subsequently remembered or forgotten events can be grouped to enquire about brain areas whose activation during item encoding becomes predictive of subsequent memory. Furthermore, fMRI data can be analysed using tools such as Statistical Parametric Mapping (SPM) (Friston et
which allow description of parametric modulations of BOLD signal as a function of time. This application is useful in learning studies where the dynamics of information encoding implicitly contain temporal changes in patterns of brain activity.

The work presented in this thesis concerns the general theme of information acquisition in various sensory modalities. Consequently, it seems pertinent to integrate notions from the field of learning and memory and multisensory processing. The term 'multisensory' is preferred in this context since it refers exclusively to multiple sensory modalities. Another similar term is 'multisensory', which can also be used to refer to the use of more than one technique, for instance. In this introduction I will review the main conceptual topics surrounding my experiments. I will present various aspects of associative learning and memory (psychological, neuroanatomical, computational, and neuropsychological), followed by a theoretical and neuroanatomical background of multisensory integration.

---

**Associative Learning and Memory**

Memory is the capacity of an organism to retain information about itself and the environment in which it lives (Fuster, 1999). The process of acquiring such information is known as learning. In this context, when dealing with learning involving information in distinct sensory modalities, I shall use the term 'crossmodal learning'. Since learning and memory are two ends of the same spectrum, I shall give an overview of the general field, starting with a taxonomy of the different types of memory and a brief review of relevant published data. Subsequently I will introduce the specifics of the type of learning investigated in the following chapters.
A Taxonomy of Memory

The basis of individual memories is facilitation or generation of synaptic connections between neurones that code for co-occurring or temporally related sensory and/or motor events, thus memory results effectively from associative processes (Fuster, 1999). Furthermore, these neuronal connections constitute cortical networks of variable distribution and proportion, including association cortices and links between association and primary cortices.

The most common criteria for classification of memory systems are: the kind of information processed; storage capacity and duration; the stages of information processing, i.e. acquisition/encoding, storage, retrieval, and manipulation/adaptation; and, the neuroanatomy underlying such operations.

1. Kinds of information

A major distinction exists between various levels of conscious engagement during memory operations involving different types of information. Cohen & Squire (1980), Cohen (1984) and Squire (1986) have used the ‘declarative versus procedural memory’ nomenclature, which implies that information can be processed consciously or unconsciously, respectively.

Declarative memory, also referred to as explicit memory, includes processing of two types of information: events and facts. Specific to events is that they are framed in time and space, such as in the case of personal episodes (Tulving, 1983, 1987). And thus, such memories are also known as episodic. In contrast facts, or semantic memory, are characterised by their acontextual nature (Tulving, 1987). There is not much evidence that the neural substrates underlying these two divisions of declarative memory are separable.

Procedural memory, also known as nondeclarative or implicit memory (Zola-Morgan & Squire, 1993) relates to various skills and mental processes, which are
preserved in amnesic patients. Such individuals are still able to learn sensorimotor
tasks (Cohen & Squire, 1980). Another type of nondeclarative memory, priming, is
also spared in amnesics (Tulving et al., 1982). Priming describes the phenomenon
of facilitated recognition of a stimulus after exposure to a different but related
stimulus. Such facilitation is independent of conscious recollection of the related
stimulus *per se* or the link between the priming and the target stimuli. Indeed,
medial temporal lobe lesions seem to impair declarative memory and leave
procedural memory unaffected. This suggests that the neural structures supporting
these two memory functions are located in different brain regions.

Finally, information acquired through the senses is known as perceptual memory,
and it can be semantic or episodic. The neural structures underlying this type of
memory include primary and association sensory cortices. Much of this thesis
involves the study of perceptual memory processes in the episodic sense, and
more specifically how perceptual information in different sensory modalities is
encoded, associated and/or integrated into unitary representations. It should be
clarified that since neuroanatomically perceptual cortex is well defined, it seems
feasible that perceptual representations conform a type of memory, which could be
defined as perceptual memory and thought of as a separate entity when compared
to motor memory, for instance.

2. Storage capacity and duration
Regarding storage capacity and duration the most traditional division is that of
short-term and long-term memory. These concepts resemble James’s (1890)
original distinction between primary and secondary memory, which later became
the basis of Waugh and Norman’s model (1965). This model describes verbal
items reaching primary memory and being retained at this stage through rehearsal,
or alternatively being forgotten, whereas rehearsed items are relayed on to
secondary memory and adjoined to permanent memory. Thus, the main distinction
between these two classifications of memory lies in their temporal persistence. The
specific retention interval estimated for short-term memory ranges from seconds to
minutes, whereas in long-term memory it can be up to years. The information capacity in short-term memory is limited, whereas in long-term memory it is not.

Evidence that this distinction in memory exists is also found in amnesic patients. Specifically patients with Korsakoff’s syndrome have long-term memory impairments whereas short-term memory is preserved. Scoville and Milner (1957) among others have also reported medial temporal lobe patients with preserved long-term memory who fail to retain new information.

This dual-system model of memory assumes that information is retained in short-term memory before it is transferred into long-term memory. A challenge to this notion has been posed by Craik (1983) who argues that long-term memory is largely dependent on the depth of elaboration during initial encoding. According to this theory only the amount of previous information elaboration can guarantee its subsequent storage, and is thus contradictory to the concept of generic transfer from short-term to long-term stores. Neuropsychological evidence also supports a view that deficits in short-term and long-term memory arise from damage to different cortical structures.

As previously mentioned, short-term memory was originally referred to as primary memory (James, 1890), and defined as the memory of the immediate past still in consciousness and the object of selective attention. In this case, the duration is modifiable by thought and rehearsal, and thus under subjective control. However, what is normally referred to as short-term memory (with a few seconds duration) is also termed immediate memory. One specific case of short-term memory is iconic memory (Coltheart, 1983), which refers to the capacity to retain a sensory item for up to one second after presentation.

A final definition of short-term memory, which by-passes to a certain degree the concept of duration, is that of working memory. Derived from cognitive psychology (Baddeley, 1983), and also termed operant or provisional memory, the concept of
working memory relates to the temporary storage used in performance of cognitive
behavioural tasks. Specific neuroanatomical claims highlight prefrontal cortex as
the main functional substrate of working memory (Fuster, 1989). The concept of
working memory is compatible with the view that short-term memory consists of
temporary retention of information for immediate use in behaviour, but not
necessarily as a gateway to long-term memory or as a storage stage previous to
consolidation.

Several instances exemplify how short-term and long-term memory are connected.
One such case is acquisition and recall of verbal material, which is determined and
facilitated by existing associations in long-term memory (Buschke, 1975).
Additionally, long-term memory can also influence acquisition and recall of non-
verbal material by contextual associations. From the neuronal point of view, it is
feasible that a network, which becomes activated by new information, may be
under the influence of existing representations. According to this view, no single
network is responsible for processing only short-term or long-term memories.

My view is that the different kinds of memory are probably based on the same
associative principles and differ in the kind of material processed, with different
neuronal populations in the brain being specialised for various types of material
(visual, auditory, verbal, etc.) and not necessarily differing in the nature of their
neural operations. Thus, procedural memory allows us to make associations
between our motor outputs and their outcomes. Episodic memory allows us to sort
out information related to events in time and relate that to previous events and to
derive experience which might be useful for future instances. Semantic memory
also associates concepts together, or words and concepts, or unites different
aspects of a fact. Even working memory has a very strong component of
comparing online information with previously stored information to produce
adaptive responses.
A related issue is also what factors lead or optimise memory retrieval. In this sense two competing theories claim that retrieval success is directly linked to the depth of processing of the material during encoding (Craik & Lockheart, 1972), whereas the other posits that the key factor is that the retrieval strategy resembles the encoding strategy (Morris et al., 1977). In my view, both these theories are related in so far as the latter proposal, transfer appropriate processing, might hold true if information is not encoded very deeply, whereas if encoding involves a considerable amount of elaboration and activation of existing associations, whether the encoding and retrieval strategies differ or not might not be crucial.

3. Psychology of Associative Learning

Associative learning has been a long-standing topic of debate in Psychology, with many theories having emerged over the years, ranging from classical conditioning to neurocomputational models, encompassing nearly every type of association. One of the most characteristic types of association is prediction or causation, i.e. how we learn that event A is predictive of event B. Knowledge about predictive relationships can determine our actions. In other words, the reason why we perform an action (A) is because we learn that it causes a desired outcome (B). But, how does such learning occur? And, how are associative relationships represented? From now on I will refer to a predictive event as ‘cue’ (c), or ‘action’ (a) if it is an external event, and the predicted event will be called ‘outcome’ (o), if it is a response.

One of the basic concepts in associative learning is that of contingency, or the temporal correlation between two events. Contingency can be measured by the probability of an outcome given an action, relative to the probability of the same outcome given no action. Sensitivity to contingency has been clearly established by Rescorla (1968) in conditioning experiments with animals, and subsequently in humans, especially in situations where there is just one action and one outcome (Alloy and Abramson, 1979). Experimental data show that if subjects’ are made to
judge between various probabilities, their performance correlates strongly with the actual degree of contingency (Wasserman et al., 1993).

However, in reality a cue or action is not always the only concurrent event, which can potentially be related to a specific outcome. Instead, multiple potential causes might be present, whereby causal selectional processes need to be in operation. A well-known phenomenon regarding this issue is ‘blocking’, which relates to the false assumption of predictive relationships between events due to previous learning of a valid predictive relationship between similar events (Chapman and Robbins, 1990; Price and Yates, 1993).

An alternative associative learning theory is the least mean square (LMS) rule (Sutton & Barto, 1981; Stone, 1986), based on the premise that the strength of a relationship between cause and outcome is determined by the weight of their connection, and where the expectation of an association can be calculated. This model also accounts for contingency effects. As an example of how this rule is applied, let us consider we have two potential causes, an action and a context/background. If the probability of an outcome given an action is greater than zero, then the larger it is, the larger the weight the background gains, since many outcomes can occur in the presence of background alone.

This model generates the correct pattern of judgements in test subjects. It also reproduces the association acquisition profiles under different contingencies (i.e. negatively-accelerated curves for highly positive and negative contingencies) and the bias in the noncontingent conditions. The LMS rule explains selectional processes, such as blocking, as a result of competition in the rule between the different potential causes of an observed effect.

Following this somewhat abstract introduction to associative learning, I will now consider the question of where in the brain these associative operations are likely to take place.
4. Neuroanatomy of Memory

With the advent of neuroimaging techniques, more detailed neuroanatomical information regarding the brain structures underlying the memory systems has become available. In addition, a good proportion of our knowledge originates from neuropsychological studies, some of which is mentioned in the previous and in the present sections. I shall focus specifically on describing the neural systems involved in episodic memory (encoding and retrieval), which constitutes the function of interest in my learning experiments.

As has previously been mentioned, encoding and retrieval are two ends of the same memory spectrum. Although some work suggests that brain structures involved in information encoding are re-activated during retrieval (e.g. Nyberg et al., 2000), some other empirical data suggests that there are basic differences between the structures underlying these two processes (Craik et al., 1996). In fact, several PET studies have focused on a comparison between encoding and retrieval as involving two distinct neural systems (Fletcher et al., 1995b; Fletcher et al., 1998a, 1998b).

In this sense, it is relevant to highlight that the intermediate process, i.e. memory storage, is considered here as patterns of activations within neocortical networks, such that memories might not lie in a specific “place” but in the various connections within a network. Therefore, perception and retrieved representations are not necessarily different in nature, but merely activations or re-activations of networks. Additionally, different aspects of a memory are likely to be coded according to a synchrony or near-synchrony principle whereby inputs that occur simultaneously or in temporal proximity are likely to be stored together. Thus, when two or more inputs have been processed simultaneous or close to each other a summation of excitation might occur allowing subsequently single inputs to produce the same neuronal response. From the neuroanatomical point of view, primary sensory areas in this context are also part of the neocortical networks mentioned above, and thus play a role at least in the initial perception (activation of a network) and processing
of various inputs which might lead to long-term memories provided these occur in some kind of associative sequence.

In relation to the encoding stage, PET studies of intentional encoding have described a network involving left prefrontal, left temporal/fusiform, and anterior cingulate cortices (see review by Cabeza & Nyberg, 1997). Specifically when subjects are asked to memorise words (intentionally) different fMRI and PET studies have shown involvement of left prefrontal cortex (Fletcher et al., 1995b; Kapur et al., 1996; Kelley et al., 1998). Everyday life information can also be encoded incidentally. Psychology experiments in which no explicit instructions are given to remember, have shown that retrieval can vary according to how information is processed at encoding (Postman, 1964; Craik & Lockhart, 1972; Hyde & Jenkins, 1973). This is exemplified by the well-known levels-of-processing effect (Craik & Lockhart, 1972).

Neuroimaging studies have shown that the same frontal regions, which activate during intentional encoding are also engaged in incidental tasks, even though for the latter the effectiveness of learning is lower. More specifically, incidental encoding of words through meaning-based judgements involves left frontal activations (Demb et al., 1995; Gabrieli et al., 1996; Kapur et al., 1994; Buckner & Koutstaal, 1998; Wagner et al., 1998a). However, when the task involves shallow processing, frontal activity is reduced, and subsequent word recollection is poor.

The frontal regions referred to here are functionally heterogeneous. The most common frontal areas of activation during encoding are located dorsally, along the inferior frontal gyrus (BA44,6) and ventrally (BA44,45,47,10) (Buckner et al., 1995a; Buckner, 1996; Petrides & Pandya, 1994; Petrides et al., 1995).

Further research has tackled the question of functional heterogeneity directly, by means of administering deep and shallow encoding conditions and observing similar activations within the dorsal part of inferior frontal gyrus for both conditions, and in the ventral part preferentially for the deep encoding condition (Logan et al.,
This finding is consistent with reports of functional dissociations within left frontal cortex (Buckner et al., 1995a; Buckner, 1996; Poldrack et al., 1999). This left prefrontal cortex is selectively involved in processing verbal material, and further divisions within it may contribute to different aspects of encoding such information. For instance, it has been suggested that left dorsal prefrontal cortex plays a role in phonological (i.e. shallow) processing of verbal material, whereas the more ventral parts of this region are recruited during semantic processing (Buckner, 1996). This might explain why activations in ventral inferior prefrontal cortex are observed as a predictor of subsequent memory, since semantic encoding generally leads to greater retrieval.

Notably, the brain areas engaged during episodic encoding are a function of the type of material presented. Thus, in contrast to the cortical regions involved in episodic verbal encoding, frontal contributions to episodic encoding of nonverbal material are likely to differ. Behavioural literature has testified to the presence of a picture superiority effect, which implies that pictures are more memorable, possibly because they can be encoded both verbally and pictorially (Paivio & Csapo, 1973; Paivio, 1986). This is a good argument for the existence of multiple codes contributing to memory. Further evidence comes from neuropsychological studies, which show that verbal and nonverbal information may be processed in different hemispheres (Milner & Taylor, 1972; Riege et al., 1980; Whitehouse, 1981; Gazzaniga & Smylie, 1983). Evidence that the right hemisphere has an advantage in processing nonverbal information comes from observation of split-brain patients who perform better on face memory tests when faces are presented to the right hemisphere than when presented to the left hemisphere (Gazzaniga & Smylie, 1983).

Other neuroimaging studies have highlighted the fact that encoding material which cannot be easily verbalised, such as unfamiliar faces or texture patterns, preferentially activates right frontal regions (Kelley et al., 1998; Wagner et al., 1998a; McDermott et al., 1999). Additionally, nonverbal encoding has been shown
to engage other cortical areas, such as parietal and occipito-temporal regions (Nyberg et al., 1996b), which are likely to contribute to the process of memory formation in various ways. For example, the initial encoding and subsequent retrieval of sounds and pictures have been observed to correlate with activations in temporal and occipital areas, respectively (Zatorre et al., 1996; Wheeler et al., 2000).

From the neuroanatomical point of view, it has been hypothesised that information reaching the neocortex converges in a central structure, which is accountable for short-term processes involving association formation between representations. Thereafter the processed information is relayed back to the neocortex for long-term storage (Marr, 1971; Squire, 1987b; McNaughton & Nadel, 1990; Gluck & Myers, 1993). This suggested structure is the hippocampus, which is located deep inside the brain in the medial temporal lobe. The hippocampus is part of the hippocampal formation, which includes the dentate gyrus and subiculum. More precisely it is an infold of the parahippocampal gyrus into the inferior temporal horn of the lateral ventricle. Out of the three structures comprised by the hippocampal formation, the hippocampus is the best studied in humans.

Some studies have identified a role for the hippocampal formation in episodic encoding (e.g. Haxby et al., 1996; Kapur et al., 1996; Stern et al., 1996). This finding coincides with the traditional view of a hippocampal role in learning of new episodes (Squire, 1992). More detailed interpretations of the role of left hippocampus in episodic memory encoding include engagement in novelty detection of verbal material (Dolan & Fletcher, 1997; Strange et al., 1999; Martin, 1999). Similarly, the right hippocampus has been implicated in novelty detection of nonverbal material (Tulving et al., 1994c; Martin, 1999). Additionally, a role for left prefrontal cortex has been suggested in relation to associative semantic processing (Dolan & Fletcher, 1997; Thompson-Schill et al., 1998). These two findings support the novelty/encoding hypothesis put forward by Tulving et al. (1996), which claims that encoding consists of two subprocesses: (1) novelty
assessment (mediated by hippocampal regions), and (2) meaning-based encoding operations (mediated by left frontal lobes). It thus follows that if information is novel, it is more likely to be further processed, in terms of a search for meaning.

Stimulation and ablation of the hippocampus can have a wide spectrum of effects, ranging from changes in behavioural and endocrine functions, attention and alertness deficits, anterograde amnesia (Squire, 1987a), and seizure activity. Anterograde amnesia is characterised by an inability to acquire new episodic information. Patients suffering from anterograde amnesia might experience some retrograde amnesia, although it normally affects recent episodes more than information acquired long ago before the trauma (Squire, 1987a). There is neuropsychological evidence suggesting that the hippocampus is involved in laying down new memories. Indirect memory tests in animals have also shown that hippocampal damage leads to learning deficits (Eichenbaum, 1992).

Remarkably, it has been observed that lesions affecting the hippocampus do not affect procedural learning in animals or humans (Cohen, 1984; Gabrieli et al., 1995). Procedural learning, however, is acquired over many trials and does not require formation of memories for single events. However, hippocampal damage does impair repetitive tasks of the conditioning type, if a certain level of complexity is present, such as comparing multiple stimuli or attending to context (Hirsh, 1974; Rudy & Sutherland, 1989, 1995). Besides, most learning deficits attributed to hippocampal damage can be due to context effects (Hirsh, 1974). For instance, amnesics with hippocampal damage might remember an event but not where or when it happened, or they might seem unaware of knowing certain information until they are indirectly prompted (Weiskrantz & Warrington, 1979; Haist et al., 1991). Similarly, animals trained to respond to a stimulus in one environment might respond less efficiently if tested in another environment (Hall & Honey, 1989). However, if an animal has hippocampal damage, the response will remain the same, no matter whether the test environment is new (Honey & Good, 1993; Penick & Solomon, 1991).
More recent fMRI studies of subsequent memory confirm the importance of medial temporal and prefrontal brain regions in the process of information encoding (Brewer et al., 1998; Fernández et al., 1998; Wagner et al., 1998b; Wagner et al., 1999). These studies show a positive correlation between the magnitude of activations of these brain regions during study and subsequent memory performance. The idea motivating these investigations originated from electrically recorded scalp potentials at the time of encoding, which revealed greater activity for words later remembered than for those subsequently forgotten (Fabiani et al., 1986; Paller et al., 1990; Rugg, 1995). The application of neuroimaging techniques to this problem, and more specifically event-related data acquisition, has allowed not only greater spatial resolution but also the capacity to identify regions of activation which correlate with learning of subsequently remembered or forgotten single study items (Rotte et al., 1998; Wagner et al., 1998b; Alkire et al., 1998; Brewer et al., 1998; Wagner et al., 1999). These subsequent memory studies were based on semantic incidental encoding tasks, followed by a recognition test which allowed labelling of studied items as remembered or forgotten. This method of separating items as remembered or forgotten was utilised in this thesis (see chapter 7).

In summary, the evidence suggests that both frontal and medial temporal regions are crucial for episodic memory formation. The way in which these two regions might interact is likely to be by frontal areas providing inputs to medial temporal areas during encoding (Moscovitch, 1992; Buckner, 1999; Buckner et al., 1999). This concept is in accord with a proposal that medial temporal cortex is involved in binding, i.e. cohesively integrating information, leading to long-term memory (Cohen & Eichenbaum, 1993; Moscovitch, 1994; Johnson & Chalfonte, 1994; McClelland et al., 1995; Schacter et al., 1998; Cohen et al., 1999).

The possibility that frontal cortex modulates medial temporal cortex during encoding is supported by lesion studies, which show that medial temporal lobe lesions leave frontal functions intact while impairing memory formation. An fMRI
study of an amnesic subject illustrates that frontal functions, such as online task goals, e.g. meaning-based elaboration, can be conceived as contributing to the process of episodic encoding, even though they are also operational when memory formation is not taking place (Buckner & Koutstaal, 1998). This might suggest that the flow of information during encoding is from frontal to medial temporal structures (Thierry et al., 2000). Hippocampo-cortical projections, however, can also be activated during memory retrieval (Schacter et al., 1998; Rolls, 2000a; Nadel et al., 2000) or novelty detection (Tulving et al., 1994c), thus suggesting that the products of medial temporal processing can have an effect on frontal processing.

Frontal contributions are essential during episodic memory encoding, as evidenced by impairments observed in frontal lesion patients (Schacter, 1987b; Shimamura et al., 1991; Wheeler et al., 1995). Therefore, not only do frontal lesions have repercussions on episodic memory encoding by impairing the input system into medial temporal structures, but they also affect online processing, verbal functions in the case of left frontal lesions (Geschwind, 1979) and visuospatial functions in the case of right frontal lesions (Corballis et al., 1999). Consequently the side of frontal lesion selectively affects episodic memory encoding by virtue of the type of online processing disturbed (Buckner & Tulving, 1995). Several studies have shown that left frontal damage selectively impairs memory for words, whereas right frontal damage impairs memory for nonverbal items (Riege et al., 1980; Whitehouse, 1981). A similar laterisation (left, right) effect related to material type (verbal, nonverbal) has also been reported for medial temporal lobe structures (Tulving et al., 1994a; Dolan & Fletcher, 1997; Strange et al., 1999; Martin, 1999; Köhler et al., 2000).

In the retrieval domain, a distinction from cognitive theories of episodic memory is between retrieval attempt and actual information retrieval or retrieval success (Tulving, 1983; Shallice, 1988; Moscovitch, 1992). PET data seem to support this distinction by providing evidence for a right prefrontal role, specifically BA10 anteriorly, during successful episodic retrieval (Buckner & Tulving, 1995;
Ungerleider, 1995; Buckner, 1996; Buckner & Petersen, 1996; Cabeza & Nyberg, 1997; Fletcher et al., 1997). Further evidence for a role of right prefrontal region in retrieval of not only items but also of temporal and spatial information (Nyberg et al., 1996a; Köhler et al., 1998), regardless of retrieval success (Kapur et al., 1995; Buckner et al., 1998), suggests that the role of this region might be specifically in adopting a retrieval mode (Nyberg et al., 1995).

Other studies show additional increased activity in anterior cingulate and cerebellum (Nyberg et al., 1995; Cabeza & Nyberg, 1997). This activation, which is stronger during recall than during recognition has been interpreted as self-initiated processing. Notably, episodic retrieval has also been observed to correlate with decreased activity in bilateral temporal regions (Nyberg, 1998).

There is evidence to support the idea that actual retrieval is mediated by brain regions which differ from those involved in adopting a retrieval mode. Successful retrieval has been associated with activations in medial temporal regions, including hippocampus (Grasby et al., 1993; Eustache et al., 1995; Cahill et al., 1996; Nyberg et al., 1996a; Schacter et al., 1996; Rugg et al., 1997; Heckers et al., 1998). This correlation between performance and brain activation seems to reflect recollective experience or confidence (Nyberg et al., 1996a; Schacter et al., 1996). This idea is consistent with the view that medial-temporal cortex acts as an initial store (McClelland et al., 1995), such that hippocampus might only play a temporary role in the formation of memories. This would explain why these activations are found at less practised stages relative to practiced stages (Petersson et al., 1997).

Therefore, the evidence so far seems to suggest that hippocampus plays an important role in the novelty detection aspect of learning and in making associations between new and existing information. This seems to be the case when the information is processed explicitly, i.e. with consciousness, rather than implicitly, such as in priming and to a certain extent in procedural learning, as well.
Successful retrieval has also been correlated with bilateral prefrontal activations (Rugg et al., 1996). To reconcile this finding with the view that prefrontal cortex is involved in retrieval mode, it could be argued that right prefrontal cortex is sensitive to retrieval intention but might be further activated if retrieval succeeds (Fletcher et al., 1997). However, it is also feasible that even within right prefrontal cortex, and consistent with findings of functional heterogeneity of prefrontal cortex (Petrides & Pandya, 1994; Buckner, 1996), different subregions respond to different degrees of retrieval success.

Notably, parietal cortex also seems to contribute to the process of retrieval. More specifically, activation in a medial parietal region, known as precuneus, has been consistently observed (e.g. Fletcher et al., 1995a). The role of this region in retrieval has been suggested to be in imagery as a mnemonic strategy (Dolan et al., 1997). Another interpretation is that precuneus activation reflects level of retrieval success (Kapur et al., 1995), possibly by means of reactivating stored representations (Roland & Gulyás, 1995). Thus the role of the precuneus in memory retrieval needs further clarification.

Generally speaking, not only medial, but also lateral parietal cortices are classified as remote visual-association areas participating in the generation of visual images of spatial scenes from memory (Roland et al., 1987). Laterally, parietal cortex has also been reported to play a role in explicit retrieval, and more specifically in verbal retrieval success (Donaldson et al., 2001). In a recognition test, left lateral parietal cortex has also been shown to respond preferentially to correct old as opposed to correct new items (Henson et al., 2000). Additionally, another study found that in an old/new episodic recognition task, hit trials (correctly recognized old items) and correct rejection trials (correctly rejected new items) evoked differential activation in left lateral parietal cortex (Konishi et al., 2000).

Lateral parietal cortex is also known for its role in decision making. In monkeys it as been demonstrated that activity of these neurons is sensitive to the probability
that a particular response will result in a gain. When animals can choose freely between two alternative responses, the choices subjects make and neuronal activation in this area are both correlated with the relative amount of gain that the animal can expect from each response (Platt & Glimcher, 1999). More specifically, the lateral intraparietal cortex of monkeys has been found to mediate performance at a visual motion discrimination task (Shadlen & Newsome, 1996). In humans, this role for lateral parietal cortex has also been instantiated. For example, a study showed that lateral parietal lobe activations are associated with shifting criteria (a manipulation of bias/decision criterion) (Miller et al., 1996).

Neuropsychologically, a case has been made for retrograde amnesia for personal material correlating with reduced perfusion in right frontal and temporal areas, whereas other areas, including lateral parietal cortex, activated normally during a PET of episodic memory retrieval (Markowitsch et al., 1997). This account poses the question of what exactly the role of lateral parietal cortex is in retrieval. Some evidence from recognition tests shows that lateral parietal lobe is associated specifically with decision making manipulations rather than with memory-retrieval processes per se (Miller et al., 2001).

Another neuroanatomical distinction has been made between areas engaged by different types of recall tasks. Generally, cued recall and recognition have been shown to involve left prefrontal and parietal activations for verbal study material. Moreover, enhanced responses have been observed in left prefrontal, left parietal and posterior cingulate for cued recall relative to recognition, and in right lateral and medial prefrontal cortex for familiarity judgements (recognition) relative to cued recall (Henson et al., 1999a). The latter activations might correlate with increased monitoring demands due to a lesser certainty of judgement.

Another type of memory described in a previous section of this introduction was priming. This type of implicit memory is expressed when repeating an item facilitates performance on a later task. The classical priming effect reported with
neuroimaging experiments is that of decreased activation in critical areas, e.g. in posterior visual areas if the task was visual, involving either words (Squire et al., 1992; Buckner et al., 1995b; Schacter et al., 1996; Backman et al., 1997; Blaxton et al., 1996) or pictures (Martin et al., 1995b). This reduction of brain activation is putatively correlated with more efficient perceptual processing, as demonstrated behaviourally by faster reaction times (Buckner et al., 1995b).

More recent studies have shown that priming effects, which need not only be perceptual, are reflected in reduced activation in sensory cortex. Conceptual priming effects have also been observed to produce reduction in the response or level of activity in high order brain areas compared to control conditions (Raichle et al., 1994; Demb et al., 1995; Gabrieli et al., 1996). For instance, decreases in left prefrontal areas have been associated with a verbal generation task requiring semantic access (Raichle et al., 1994) and with a semantic decision task (Demb et al., 1995; Gabrieli et al., 1996). Such priming effects seem to be specific to situations where both item and task are repeated (Wagner et al., 1997). Generally, priming effects seem to follow the transfer appropriate processing principle (Roediger et al., 1989, Blaxton 1989; Morris et al., 1977) where facilitation across tasks occurs when the same kind of processing takes place across item repetitions.

Following this review of neuroanatomical descriptions of general episodic memory provided by neuropsychological neuroimaging studies, it is necessary to take a closer look at the ideas regarding crossmodal integration relevant to the experiments presented in this thesis to characterise the human functional neuroanatomy underlying associative memory and more specifically crossmodal association formation.
Multisensory Integration

Our perception of the world is not divided into independent sensory experiences, following inputs in various sensory modalities. Instead our experience of the world is multisensory, providing us with multiple and simultaneous sensory inputs, which we need to generate appropriate responses to the environment.

From an evolutionary point of view, the capacity to integrate information acquired through different sensory channels is highly adaptive. It allows us to attend to more than one event concurrently or to guide and enhance the focus of our attention by coherently perceiving the same object in more than one sensory modality. For instance, it is known that seeing the lips of an interlocutor while hearing their speech provides more efficient understanding of the message than if the perceived modality were only auditory (Sumby and Pollack, 1954). In our everyday lives we are constantly exposed to information delivered multisensoryly. A good example is information conveyed through the media. From this point of view, television is a more informative means than unimodal radio, for instance, by virtue of providing a visual context to the auditory inputs.

The origins of evolutionary advantage to multisensory perception are probably in the natural world. There are many instances when having knowledge of different sensory attributes in an object can be profitable. For example, primates learn that fruits of different colours have different gustatory qualities, such that intense red fruit is associated with ripe sweetness, which is ultimately correlated with greater nutritional value (Murray et al., 1998). Of relevance is also the capacity to make associations through repeated exposure to certain smells with various degrees of edibility, which is necessary to avoid consumption of potentially damaging foods. In other cases, learnt associations of various perceptual features of an object can provide the possibility of accessing information through alternate channels when
access to one sense is denied. For instance, at night, auditory information might be sufficient for an animal to recognise the presence of a predator.

Another valuable feature of crossmodal associations is learnt predictiveness. In many situations, access to information in one sensory modality might not be denied categorically but temporally. For example, the sound of a nearing predator in bright daylight might be predictive of its nearing presence. Upon arrival the expectation created by one modality (auditory) is confirmed by simultaneous multisensory perception (auditory and visual), or by matching the first stimulus to an associated stimulus in a different modality (auditory then visual). Thus, sensory synergy across various modalities conveying matching or complementary information can result in an overall increase in reaction speed (Hershenson, 1962; Morrell 1968a,b, 1972; Bernstein et al., 1969; Andreassi & Greco, 1975; Posner et al., 1976; Gielen et al., 1983; Colavita, 1974; Colavita & Weisberg, 1979).

In this respect, single modalities have different effects on reaction speed. For instance, due to relative differences in processing time in the inner ear and the retina, auditory stimuli allow faster (40-60 ms) responses than visual stimuli. Remarkably, when a visual stimulus precedes an auditory stimulus, their difference in processing time results in simultaneous perception, and in turn in increased amplitude of the evoked potentials to the combined stimuli (Andreassi & Greco, 1975; Costin et al., 1991). This increased magnitude in evoked potentials might reflect greater salience of the stimulus, which is thus rendered less ambiguous, allowing for faster responses from the perceiver. Evidence that this is the case comes from studies of mentally retarded and dyslexic children who experience more difficulty dealing with multisensory than unimodal inputs, leading to patterns of evoked potentials with less multisensory summation than those observed in normal children (Shipley, 1980).

One could argue, however, that vision preceding audition is not a common occurrence. Maybe reading aloud is a case where visual processing precedes
auditory processing, which subsequently leads to audiovisual integration. However, rather than through consecutive presentations, event potentials are known to be amplified during simultaneous multisensory presentations (Foxe et al., 2000; Calvert et al., 2000).

A possible mechanism for this type of synergy might lie in a summation of the relative energies from both stimuli (Nickerson, 1973). In fact, there is evidence that certain multisensory neurones respond more efficiently to multisensory stimulation than would be expected by summing the responses to individual stimuli. This phenomenon is known as response enhancement (Stein & Meredith, 1994; Wilkinson et al., 1996; Stein, 1998). An alternative mechanism for sensory synergy might rely on learnt associations across modalities producing a preparation enhancement, similar to an alerting effect of one stimulus on the other (Nickerson, 1973). A combination of both these mechanisms is another possibility (Bernstein, 1970; Welch & Warren, 1986).

Whether multisensory integration and crossmodal attention are intrinsically linked processes remains controversial. There is some evidence that crossmodal integration and crossmodal attention function under different mechanisms. For instance, multisensory integration occurs in anaesthetised animals, i.e. without intent (Jiang et al. 2001). Also, the ventriloquism effect, based on multisensory integration, occurs pre-attentively and independently of voluntary and involuntary spatial attentional shifts (Bertelson et al., 2000; Vroomen et al., 2001). Furthermore, approximate temporal synchrony is required for multisensory integration but not for involuntary spatial attention effects to take place even if stimuli are brief (<100 ms) and separated by 100 to 500 ms (McDonald et al., 2000). Multisensory integration is greatly reduced under such conditions (Meredith et al., 1987; Jack & Thurlow, 1973). These studies seem to suggest that involuntary shifts of attention are produced by stimulus-driven processes independent from multisensory integration. On the other hand, other findings show that modality convergence patterns, sensory response properties, and principles
governing multisensory integration in the superior colliculus in alert and anaesthetised animals are similar, thus suggesting that there might indeed be a functional link between multisensory integration and cross-modality attentive and orienting behaviours (Wallace et al., 1998).

Returning to examples of multisensory integration, a good case is the common occurrence of uncertain perception of a familiar voice in the context of a noisy environment, which is only confirmed with simultaneous visual access to the corresponding face. In fact, it has been shown that access to the speaker’s lip movements can alter speech comprehension in a manner equivalent to altering the signal to noise ratio of the auditory stimulus by 15-20dB (Sumby & Pollack, 1954). In other situations, where accessing one sensory modality can lead to compensation from an inaccessible modality, constancy of multisensory identities is crucial, i.e. learnt coherence between associated inputs in different modalities, e.g. face-voice. Several studies have actually shown that providing information in one sensory modality can activate brain areas that are normally engaged in processing equivalent information in a different modality. For instance, lip reading causes activation in auditory cortex (Sams et al., 1991) even if lip movement is presented in silence (Calvert et al., 1997; MacSweeney et al., 2000).

More evidence that learnt coherence between stimuli in different modalities can affect our perception is provided by experiments with combined nonmatching visual and auditory speech cues. Such a procedure results in an audiovisual illusion, known as the McGurk effect (McGurk & MacDonald, 1976). An example of this effect occurs when hearing “ba-ba” and seeing the mouth articulate “ga-ga”, results in the auditory perception of “da-da”. Another trick can be played to our senses, by making a dummy’s mouth move to the sound of a human voice, and perceiving the voice as coming from the dummy. The ventriloquism effect (Howard & Templeton, 1966) covers a broad range of intersensory bias phenomena, where vision can influence not only judgements on audition but also on proprioception, proprioception can bias audition, etc. (Held, 1955; Pick et al., 1969; Thurlow &
Crossmodal illusions can, however, have a dramatic impact on individuals who, on a regular basis, are subjected to compelling sensory perceptions in one modality triggered by stimuli in another modality, the synaesthetes. The word 'synaesthesia' comes from the Greek "joining of the senses". In more scientific terms, it has been defined as an involuntary phenomenon in which real information in one sense is accompanied by perception in another sense (Cytowic, 1989). This multisensory phenomenon has been known for centuries. Some of the earliest reports this century were on sonogenic synaesthesia, i.e. the induction of visual or tactile perception by music (Critchley, 1977; Henson, 1977). Research in this area is complicated by a lack of empirical testability of such subjective experiences. Synaesthesia has been explained in terms of learned associations, whereby repeated pairings of concurrent stimuli can create a perceptual synthesis (Calkins, 1893; Claparède, 1903; Dresslar, 1903; Harris, 1908). The underlying neural mechanism might consist of independent groups of neurones, which are activated in close temporal proximity, leading to concurrent activity as a consequence of repeated exposure to specific stimuli, producing the synthetic effect. However, it has been hypothesised that exposing normal subjects to learning crossmodal associations of the kind experienced by synaesthetes, would not evoke synaesthesia (Gray et al., ongoing experiment in 1997, unpublished). This hypothesis has not yet been tested.

It is also known that synaesthesia is part of the hallucinatory symptomatology of schizophrenia (Bailey & Johnson, 1997) and that synaesthetic experiences can be provoked by consumption of hallucinogenic drugs, such as LSD and mescaline (Motluk, 1994). These two cases suggest the possibility that synaesthesia might not be learnt but caused by alterations in brain processes either at the structural level, as a consequence of genetic or developmental abnormalities, or at the functional level, by exerting an artificial imbalance in the neurotransmitter systems.
An alternative and all-encompassing view is that synaesthesia can be caused by different factors.

Another property of crossmodal associations is the ability they confer to access unavailable information to one sensory modality, through imagery, triggered by presentation of a cue in an alternate modality. This is the case when hearing a familiar sound reminds us of an associated visual attribute, such as when hearing a cow moo evokes a mental image of the animal. The associated mechanisms are very much memory-based and combine a bottom-up input of the triggering stimulus with the top-down retrieval of associated information. In psychological terms, this phenomenon is also referred to as crossmodal matching or crossmodal recognition, and can be artificially probed by presenting one of two objects as a sample in one modality and having to select the same object when the two are presented together for choice in an alternative modality.

There are several versions of crossmodal matching (Bryant, 1968), one consisting of a simple same-different judgement, whereby only one stimulus is presented in the second modality (the stimulus being the same as or different from the stimulus presented in the first modality). In another version, one stimulus is presented in one modality, and then has to be recognised from among a set of stimuli presented in the other modality. The most common task used in crossmodal matching studies consists of presenting an object visually and later providing somatosensory cues to judge whether the item corresponds to the one previously seen. In apes, crossmodal matching has been demonstrated bidirectionally in the visual and tactile modalities (Davenport & Rogers, 1970; Davenport et al., 1972; Davenport et al., 1973; Davenport et al., 1975). Such a task is also easily performed by normal human adults and even young children (Abravanel, 1981; Hadjikhani & Roland, 1998). In the visual and auditory modalities, match incongruence can also be effortlessly detected (Lyons-Ruth, 1977). Another training procedure to test crossmodal performance is known as transfer, and consists of training subjects to discriminate two objects in one sensory modality and then presenting them with the
same stimuli for discrimination in another sensory modality. If subjects learn faster in the second modality than a control group whose experience in the first modality was with a different pair of stimuli, then crossmodal transfer has occurred (Murray et al., 1998).

Regarding invoking representations through a deprived modality, it is relevant to discuss the case of sensory loss patients, where deprivation of one sensory modality is not contextual but organic. In this case, crossmodal associative learning can compensate for the sensory loss by replacing the information entry through the impaired sensory channel by an alternative access modality. The best known examples are those of blind Braille readers and deaf lipreaders.

It has been demonstrated that visual cortex in congenitally and late-onset blind patients can respond to sensory input from a different modality, such as tactile (e.g. Braille) and auditory stimulation (Wanet-Defalque et al., 1988; Veraart et al., 1990; Rösler et al., 1993; Kujala et al., 1992, 1995, 1997; Sadato et al., 1996; Weeks et al., 2000). Additionally, transcranial magnetic stimulation (TMS) of the occipital cortex in early-onset blind subjects has been shown to induce errors in identification of Braille characters and embossed Roman letters (Cohen et al., 1997). Similarly, visual stimulation through silent lipreading can evoke auditory cortex activation in normal individuals (Calvert et al., 1997; Puce et al., 1998; MacSweeney et al., 2000). It seems that lipreading improves speech perception by enhancing activity in primary auditory regions. In normal subjects this might be due to integration or substitution of the two sensory streams in heteromodal regions around superior temporal sulcus (Calvert et al., 1997). In the case of deaf long-term cochlear-implant users the interaction of listening to words and watching sign language also activates auditory cortex (Nishimura et al., 2000).

All these examples highlight the fact that making associations across sensory modalities might be one of the most basic types of associative learning and one that requires no conscious effort. Undoubtedly this capacity has afforded humans,
and many other animal species, the possibility of acquiring a high degree of adaptability to a multisensory environment. It is feasible that regardless of what combination of sensory modalities might underlie any type of crossmodal processing, the neuronal computations involved be similar. There is some evidence to suggest that crossmodal processes such as crossmodal matching and integration, for example, are different (Stein & Meredith, 1993; Radeau, 1994; Calvert et al., 1998. The difference, however, might lie on the kind of operations performed and not necessarily on the sensory modalities involved.

It is notable that the study of perception has so long been limited to investigation of individual sensory modalities. Simple observation of brain structures and the way they function and cooperate with each other should render it evident that multisensory integration is a common neural function engaging our brains.

Recently, it has been demonstrated that neuroimaging techniques not only offer a means to describe ‘where’ in the brain multisensory integration takes place but they might also inform about ‘how’ this process comes about. For instance, in addition to previous knowledge regarding the role of auditory cortex in visual speech (Sams et al., 1991; Calvert et al., 1997), a fMRI study has shown that when the speaker can be seen and heard, activation in sensory-specific visual and auditory cortices is enhanced compared to the response to either modality alone (Calvert et al., 1999). Increased activity in these modality-specific areas might be the physiological correlate of the subjective improvements in ‘hearing’ when a speaker’s lip and mouth movements are visible (Reisberg, 1987) and superior ‘visual localization’ of the sound source (i.e. to discriminate who is saying what, in a room full of speakers) when the auditory and visual speech patterns can be matched (Driver, 1996; McDonald et al., 2000).

These findings have been complemented by a further fMRI study designed to detect brain areas involved in synthesizing auditory and visual speech signals (Calvert et al., 2000), which showed that multisensory integration can be
detectable at the macroscopic level as changes in the BOLD response. This study found that semantically concordant or audiovisual speech not only revealed co-responsive brain areas, i.e. stimulated by each modality in isolation but a positive interaction effect in response to congruent audiovisual speech, i.e. a greater response to congruent audiovisual speech than to each modality in isolation. Similar results have also been reported using MEG (Raij et al., 2001). Also, an ERP study involving non-linguistic information showed evidence of multisensory integration through the enhancement principle (Giard & Peronnet, 1999). It can be thus concluded that anything less than a superadditive interaction could simply reflect the linear summation of responses from two sets of sensory-specific neurons rather than multisensory integration responses (Calvert, 2001).

However, the neural mechanisms underlying crossmodal perception are not fully understood. There are two main views regarding the issue of how inputs from different modalities are integrated. One is that stimuli are processed unimodally and independently from one another but remain accessible to other neural populations, “relay points”, which are capable of combining these stimuli to produce multisensory integration (Wallace & Averback, 1955; Ettlinger & Wilson, 1990). The other view is that sensory information is coded amodally, i.e. stimuli in all modalities are amenable to ranking following some continuous scale (number, size, intensity, duration, etc.). The scale could be the factor that allows stimuli in different modalities to be judged as equivalent. This implies that crossmodal transfer of scales such as intensity and pulse might be handled by common neural mechanisms for different sensory modalities (Sutherland, 1959; Stevens, 1975; Price 1988; Maunsell, 1989), including multisensory cells (Bental et al., 1968; Horn 1965; Spinelli et al., 1968). There is evidence that interaction between the auditory and visual pathways gives rise to convergence of flash and sound stimuli at the single-cell level in the primary visual cortex (Bental et al., 1968). Additionally, single-cell recordings in higher visual cortex have shown that some of these neurones are tuned to responding to line orientation, regardless of modality of presentation, visual or tactile (Maunsell et al., 1989). This suggests that information
coded in these neurones is amodal, but does not explain how stimuli in different modalities become related, or equivalent, to one another on some scale. More recently, neuroimaging work has provided some support for the idea that sensory-specific areas can represent multisensory stimuli (Hoerster et al., 1989; Calvert et al., 1998; Banati et al., 2000). A neuroanatomical description of brain regions putatively involved in crossmodal processes might shed some light on how information is integrated across sensory modalities.

The terminology used regarding this topic will avoid the use of ambiguous terms, such as “multimodal”, which not only refers to various sensory modalities but also to the combination of different techniques. To avoid this problem I use the terms “unimodal”/”unisensory”, “multisensory”/”crossmodal” to refer to processes involving one or more senses, respectively, since these are well represented in the literature and do not seem to lead to ambiguous semantic interpretations. Multisensory integration refers to a specific output response – the modulation of a response to one sensory modality by information in another sensory modality. Convergence of afferents from different sensory modalities in some brain areas does not necessarily imply that multisensory integration is achieved in these areas. Thus, I have referred here to multisensory cells which respond to stimuli in different modalities and to “heteromodal” or “polymodal” areas where information from different sensory modalities is integrated.

**Neuroanatomy of Multisensory Integration**

In the human brain, the auditory, visual and somesthetic association areas are relatively interconnected, with abundance of cortico-cortical connections among projection areas (Geschwind, 1965; Hewes, 1973; Jensen, 1971; Lancaster, 1968). Such connections are presumed to be essential for crossmodal functions (Myers, 1967). Most detailed neuroanatomical studies, and thus most of the knowledge we possess about connectivity in the primate brain, however, are based on nonhuman
primates. In the following sections the most relevant of these studies will be reviewed.

**Unisensory vs Multisensory Areas**

Sensory information accesses the brain through the eyes, ears, nose, tongue, and skin. With the notable exception of olfaction, sensory information reaches sensory cortex via the thalamus. From the thalamus, connections go straight into primary sensory cortices. Each of these primary sensory cortices has a specific location in the brain (Fig.1). Primary somatosensory cortex lies in the postcentral gyrus in the parietal lobe. Primary visual cortex is situated in the calcarine gyrus of the occipital lobe. Primary auditory cortex is located in transverse temporal gyri of Heschl. Primary gustatory cortex can be found in the most ventral part of the postcentral gyrus of the parietal lobe, and primary olfactory cortex in the piriform and periamygdaloid regions of the temporal lobe. The primary vestibular area is putatively located in posterior insular cortex in humans.

![Map of human cerebral cortex showing major functional areas including primary sensory cortices](Lassen et al.,1978).
Receptive fields in primary sensory cortex neurones are considered to be restricted to responding to one modality, although several studies to date have demonstrated this not to be always the case. For instance, primary visual cortical neurones can respond to auditory stimulation (Bental et al., 1968; Spinelli et al., 1968). Also, plastic changes in primary auditory cortex have been induced in the healthy brain by means of a ventriloquist's paradigm (mismatching sounds and their visual corresponding locations) (Recanzone, 1998). Additionally, EEG and MEG recordings of the human brain have detected responses in visual and auditory cortices to both visual and auditory stimuli (Schurmann et al., 1997).

There is increasing evidence to believe that what used to be considered unisensory cortex might be of crucial importance in multisensory integration. A recent fMRI study showed convergence of somatosensory and auditory inputs within in auditory cortex resulting from simultaneous stimulation in both modalities (Foxe et al., 2002). Similar results had previously been observed in awake monkeys where intracranial recordings revealed direct somatosensory inputs to the caudomedial belt area of auditory association cortex (Schroeder & Foxe, 2002; Schroeder et al., 2001). This area, which is adjacent to primary auditory cortex and receives early feedforward inputs from both the auditory and somatosensory systems, might be the monkey homologue of that reported by Foxe et al. in humans. Another recent high-density electrical mapping study demonstrated auditory-somatosensory integration in auditory cortex at relatively early latencies in humans, with suggested generators in early auditory and somatosensory cortices (Foxe et al., 2000). However, despite the data supporting early feedforward multisensory integration, further investigations are needed in order to clarify the functional significance of this early mechanism.

Another proposal is that primary sensory neurones have the capacity to respond to certain amodal (Sutherland, 1959; Stevens, 1975; Price 1988; Maunsell, 1989) features which might be common across sensory modalities, and that they respond
preferentially to one modality, unless adaptive capacity is induced under certain circumstances. Regarding adaptation, it has been shown that neonatal diversion of retinal axons to the auditory thalamus results in primary auditory cortex resembling visual cortex in its response properties and topography (Sur et al., 1990; Roe et al., 1990, 1992; Gao & Pallas, 1999; Sharma et al., 2000; vonMelchner et al., 2000), thus illustrating a great potential for plasticity in primary sensory cortices.

Damage to primary sensory areas has the most devastating effects on perception, although plastic reorganisation allows for partial recovery of function or functional substitution by structurally similar areas. In the case of the somatosensory system, ablation of the postcentral gyrus (areas 1-3 of Brodmann, BA1-3) causes complete loss of discriminative touch and proprioception and crude awareness of pain, temperature, and light touch. Lesions to the calcarine cortex (BA17) produce blindness. Depending on the site of the lesion, blindness can be partial. For instance, lesions involving the inferior calcarine cortex produce upper contralateral quadrantanopsia, i.e. blindness to the contralateral upper quadrant of the visual field. Similarly, lesions involving the upper calcarine cortex cause blindness to the contralateral lower quadrant of the visual field. Stimulation of the primary auditory cortex, in the posteriomedial part of Heschl transverse gyri (BA 41-42), produces crude auditory sensations such as buzzing, humming, or knocking, generally termed tinnitus. Lesions to this region of auditory cortex result in impairment of sound localisation in space and diminution of hearing bilaterally, but mostly contralaterally. In relation to the chemical senses, it is notable that irritative lesions to gustatory cortex (BA43), located in the parietal operculum, ventral to primary somatosensory cortex, give rise to hallucinations of taste, and ablation produces impairment of taste contralaterally. Irritative lesions to primary olfactory cortex, normally concurrent with lesions to the uncus, also give rise to hallucinations of smell. In addition, lesions to primary vestibular cortex in the posterior insular region impair perceptual judgements on body orientation and movement (Afifi & Bergman, 1998).
Adjacent to primary sensory areas are secondary sensory areas. In the case of the somatosensory system, recordings of evoked potentials show that secondary sensory areas are smaller in size than primary areas, and are generally considered unimodal. In the case of somatosensory cortex, lesions to secondary cortex (SII/SA2), produce asymbolia for pain, i.e. the absence of psychic reaction to painful sensations (Schilder & Stengel, 1938). The differentiation between secondary and association sensory cortices is not always clear. For example, in the visual system, areas BA18-19 are arbitrarily termed secondary or association visual areas. A more histology-based terminology in this system is that of 'striate' (primary) and 'extrastriate' (association) cortices. Likewise in the auditory system, neighbouring primary auditory cortex lies association auditory cortex (BA22, BA24) which is involved in more complex sound analysis, such as processing of the phonetic components of speech.

At the level of the association cortex, different properties within a sensory modality might interact. It is generally association cortices that project to multisensoryl cortices, where inputs from various sensory modalities converge. For instance, somatosensory association cortex, encompassing areas BA5 and BA7 in the superior parietal lobe, is involved in integration of various somatosensory inputs, including the perception of shape, texture, size and object identification (stereognosis).

From the neuropsychological point of view, the functions of each region can partly be distinguished by examination of the deficits caused by lesions restricted to a specific area. For instance, lesions to areas within visual association cortex can affect perception of different visual properties, according to the location of the lesion. Lesions in the region of lingual and fusiform gyri, V4, can cause achromatopsia, or loss of colour perception (Zeki, 1990; Heywood et al., 1995). Similarly, lesion to V5 is associated with akinetopsia or impaired visual motion perception (Zeki, 1991; Silverman et al., 1995; Heywood & Zihl, 1999). Areas 18 and 19 project to posterior parietal cortex (BA7) and inferotemporal cortex (BA20-
21, BA37) and, in fact, visual association cortex extends into the temporal and parietal lobes. Integration of visual information proceeding from visual association cortex into these other areas allows for more complex visual functions, such as depth perception (stereopsis), and face and object recognition. Similarly, damage to the auditory association cortex produces impairments in speech comprehension, such as in the case of Wernicke's aphasia (Wernicke, 1874; Naeser et al., 1987), and in auditory imagery (Halpern and Zatorre, 1999), but not in simple sound analysis.

Traditionally, neuroanatomical foundations regarding sensory cortices refer to a hierarchical organisation, such that relatively raw sensory inputs at the primary cortices are followed by successive stages of intramodal elaboration, allowing for more complex discrimination of stimulus features. This information is then relayed onto multisensory regions for integration (Pandya and Seltzer, 1982b). According to hierarchical models, sensory signals are elaborated at successive stages in sensory association cortices, information flows mainly unidirectionally, and some convergence takes place at each stage to allow gradual integration in multisensory areas, such as posterior parietal (Mesulam et al., 1977), superior temporal sulcus (Bruce et al., 1981), and prefrontal cortex (Bignall and Imbert, 1969; Nauta, 1971; Pandya and Kuypers, 1969).

However, an alternative approach is based on the idea that cortical functions are distributed in several parallel systems (Goldman-Rakic, 1988). This parallel distributed processing model finds support in neuroimaging studies, which show that performing a perceptual or cognitive task can increase blood flow simultaneously in multiple brain areas, which might have been traditionally considered as different stages of a hierarchy. However, it must be said that the temporal resolution of most neuroimaging techniques, with the exception of EEG and MEG, is not high enough to reliably support this claim. Further evidence comes from electrophysiological studies of the visual system, which show that different features of the visual world are processed in parallel (Hubel and...
Livingston, 1985; Shipp and Zeki, 1985). This model would also explain how segregation of input is maintained in the association networks, so as to facilitate a final bridging of sensory and executive processes. According to this model, integration across cortical networks might be mediated by local cortico-cortical connections between, for instance, subdivisions of posterior parietal (Seltzer and Pandya, 1986) or prefrontal (Barbas and Mesulam, 1981, 1985) cortices, or by multiple innervation of all components of a network by a thalamic nucleus. Regarding this last possibility, it is worth mentioning that in primates, the medial pulvinar nucleus projects to a system of cortical areas that are interconnected, namely posterior parietal, prefrontal, anterior cingulate, superior temporal sulcus, and other areas of cortex.

In conclusion, it seems that the neuroanatomical principles of segregation and integration can co-exist, such that distributed brain subdivisions, working in parallel on specific processes, are highly interconnected. Several developmental findings also support this view. Namely, synaptogenesis occurs at the same rate and reaches peak values at the same age in areas of sensory, motor, limbic, and association cortex, illustrating a high degree of integration in maturational sequence (Rakic et al., 1986). However, some challenging questions still remain to be answered, such as how the brain organises its subsystems to produce integrated behaviour.

**Multisensory association areas**

The aforementioned cortical regions, which receive input from more than one modality, are the so-called multisensory association areas. In the cortex, multisensory regions are located mainly posterior to the central sulcus (post-Rolandic), around the limbic system (paralimbic), and in the frontal lobe (Fig. 2).
Fig. 2. Diagrams showing the locations of multisensory sensory convergence areas in pre- and post-Rolandic regions on the medial, lateral, and ventral surfaces in the monkey brain (Pandya & Yeterian, 1985).

**Post-Rolandic Multisensory areas**

Within the parietotemporal region several multisensory areas have been described, namely the intraparietal sulcus (IPS), the caudal inferior parietal lobule (IPL), the caudal superior temporal gyrus, the medial parietal region, and superior temporal sulcus (STS) (Fig. 3).
Seltzer and Pandya (1980) have demarcated an area in the mid-lower bank of IPS according to architecture and connections, termed Poa, which receives input from somatosensory, visual, and vestibular areas. This multisensory region is thus involved in integrating sensory information about head position with peripheral visual input (Fuster, 1985). More specifically, it receives input from (1) first-order somatosensory association area (SA1, containing head, face and neck representations), (2) first-order visual association area (VA1) of preoccipital gyrus (Kuypers et al., 1965; Rockland and Pandya, 1981; Ungerleider and Mishkin, 1982; van Essen and Maunsell, 1983), (3) rostral IPS (Fredrickson et al., 1966; Büttner
and Lang, 1979), and (4) it has reciprocal connections with the frontal eye field region (Pandya et al., 1971; Mesulam et al., 1977; Barbas and Mesulam, 1981).

Caudal IPL (angular gyrus being the human homologue) is considered a multisensory region allowing integration between higher-order somatosensory, visual and limbic information (Mesulam, 1983). It contains two distinct architectonic areas (caudal PG and Opt), receiving inputs from (1) both somatosensory association areas, SA3 of SPL and IPL, (2) visual association area VA1 (Kuypers et al., 1965; Pandya and Seltzer, 1982a; Ungerleider and Mishkin, 1982), and paralimbic inputs from (3) cingulate gyrus (area 23) and (4) the parahippocampal region (Divac et al., 1977; Mesulam et al., 1977; Pandya et al., 1981; van Hoesen, 1982; Rosene and Pandya, 1983). It projects to other multisensory areas of STS, paralimbic regions (cingulate and parahippocampal), and prefrontal cortex (Jones and Powell, 1970; Chavis and Pandya, 1976; Seltzer and Pandya, 1976; Petrides and Pandya, 1983). Such connectivity in the caudal IPL region may underlie ecologically relevant functions, such as visuospatial attention and orientation. Several studies have showed such a role for caudal IPL in monkeys and for the human homologue, angular gyrus (Critchley, 1953; Denny-Brown and Chambers, 1958; Heilman et al., 1970; Mountcastle et al., 1975; Robinson and Goldberg, 1978; Heilman, 1979; Lynch, 1980; Sakata et al., 1981; Hyvärinen, 1982; Mesulam, 1983).

Also in the parietotemporal junction, area Tpt receives input from somatosensory and auditory association areas (Pandya and Kuypers, 1969; Jones and Powell, 1970) projecting to dorsal premotor cortex in the frontal lobe (Jones and Powell, 1970; Chavis and Pandya, 1976) and to posterior cingulate gyrus (Pandya et al., 1969). By virtue of its connectivity, Tpt might be involved in integration of auditory and somatosensory information in the face, head, and neck surfaces. This region probably plays a role in audiospatial attention as has been suggested (Hyvärinen, 1982), by virtue of its connections with cingulate and premotor regions, possibly playing a role in head turning toward significant auditory stimuli. This function in
sound source localisation contrasts with and complements analysis of spectral and
tonotopic properties carried out in primary auditory cortex (Merzenich and Brugge,
1973).

Below the cingulate sulcus in the medial parietal cortex lies an auditory-
somatosensory multisensory region, which receives inputs from (1) auditory
association area AA1 (Pandya et al., 1969) and (2) somatosensory SPL areas,
SA1 and SA2 (Pandya and Seltzer, 1982a). The connectivity of this region
suggests a role in orientation of the trunk toward sound sources.

Finally, STS contains subregions of bimodal and trimodal convergence of sensory
input (Seltzer and Pandya, 1978) (Fig. 4a). Each sensory association area has a
unimodal projection zone in STS (i.e. auditory association areas project to TAA,
visual association areas project to TEE, and somatosensory association areas
project to IPa and PGA), surrounding multisensory areas of STS, TPO and PGA
(Fig. 4b). TPO is situated in the upper bank of STS, and receives inputs from (1)
auditory association areas AA1, AA2, and AA3, (2) somatosensory association
area SA3, and (3) visual association areas VA2 and VA3. PGA is located in the
depth of STS, and receives inputs from sensory association areas, too, but is
predominantly linked to the somatosensory modality.

Multisensory properties of STS have been widely reported (Heilman et al., 1970;
Petrides and Iversen, 1978; Bruce et al., 1981; Leonard et al., 1983). Most of the
inputs to STS are from (1) third-order sensory association areas, (2) from other
polymodal areas, such as caudal IPL (Pandya and Kuypers, 1969; Jones and
Powells, 1970), (3) from prefrontal cortex (Nauta, 1964; Jones and Powell, 1970;
Pandya et al., 1971), and (4) from paralimbic regions (Pandya et al., 1981; Van
Hoesen, 1982; Amaral et al., 1983). This is suggestive of underlying sensory
integration functions of the highest level. Recent studies have highlighted the
multisensory nature of this region. For instance, a PET experiment has shown that
STS activates under conditions of selective visual and auditory attention
(Kawashima et al., 1999). This area has also been found to be relevant in processing visual motion (Ahlfors et al., 1999; Lauwers et al., 2000), mouth movements (Puce et al., 1998), and in recognising voices (Belin et al., 2000), faces and animals (Chao et al., 1999b).

![Diagram showing the architectonic parcellation of the superior temporal sulcus and surrounding cortex in the monkey brain according to Seltzer and Pandya (1978).](image)

**Paralimbic multisensory areas**

The caudal part of parahippocampal gyrus has been considered a multisensory region because it receives inputs from sensory association areas (Jones and Powell, 1970; Seltzer and Pandya, 1976), namely from (1) visual association areas
(VA1 and VA2), from (2) auditory association areas (AA2), and (3) indirectly via caudal IPL from somatosensory cortex (Seltzer and Pandya, 1984) (Fig. 5a). Within parahippocampal gyrus there are bimodal and trimodal convergence zones. Additionally, this area receives input from other paralimbic regions, such as cingulate gyrus, and from caudal orbitofrontal cortex, and connects to hippocampus via entorhinal cortex (van Hoesen et al., 1972; Baleydier and Mauguiere, 1980; Pandya et al., 1981). It has been suggested that multisensory sensory and limbic connections in this region underlie learning and memory functions (van Hoesen, 1982).

**Frontal lobe multisensory areas**

Convergence of sensory inputs also takes place in certain regions of the frontal lobe (Pandya and Kuypers, 1969; Jones and Powell, 1970; Chavis and Pandya, 1976) (Fig. 5b). Around the arcuate sulcus in the premotor region lies a site of sensory convergence from first-order association areas, with AA1 and SA1 of SPL sending inputs to the dorsal periarculate region. Also, in the central portion of the arcuate sulcus, trimodal convergence of all first-order sensory association areas takes place. Single-unit recordings have revealed polysensory (visual, auditory and somatosensory) characteristics in neurones of the periarculate region, in the context of behavioural tasks, contingent upon integration of inputs from more than one sensory modality (Welch and Stuteville, 1958; Petrides and Iversen, 1976; van Hoesen et al., 1980).

Additionally, ventral prefrontal cortex also receives bimodal and trimodal convergence, but from second-order sensory association areas. For instance, ventral area 46 receives inputs from AA2, VA2 and SA2. This periarcular convergence zone has also been implicated in behaviour involving synthesis of information from more than one sensory modality (Passingham, 1972; Passingham and Ettlinger, 1972). The human homologue of this area corresponds to the ventral part of area 9/46. In humans, it has been suggested that ventrolateral prefrontal cortex plays a role in active encoding and retrieval of specific information held in
visual, auditory and somatosensory association areas, to which it holds connections, thus allowing for selection, comparison and decision processes regarding information held in short- and long-term memory (Petrides 1994, 1995) as well as in associative learning (Passingham et al., 2000; Wishaw et al., 1992). In the monkey, lesions to this perirhinal area have been shown to cause deficits in conditional associative learning (Petrides, 1985a). More specifically, damage to this region seems to impair rhesus monkeys on a tactile-visual cross-modal matching task (Petrides & Iversen, 1976).
Fig. 5. (A) Diagram showing the location and connections of the multisensory area of the parahippocampal gyrus. (B) Diagram showing the location and connections of premotor and prefrontal multisensory areas in the monkey brain (Pandya and Yeteran, 1985).
**Insula-Clastrum**

The connectivity of the clastrum, which has been largely studied in cats and nonhuman primates, is extensive, covering virtually all of the cerebral cortex, receiving and giving rise to cortical projections, and containing maps of different sensory (visual, auditory, and somatosensory) and motor systems (Sherk, 1986).

The clastrum is a thin structure located medially to and all along the insula (Fig. 6). In the monkey, portions of the clastrum connected with TEO and TE (inferior temporal cortex) appear to overlap portions connected with other cortical areas, including V1, V2, V4, MT, and MST (visual occipital areas) (Webster et al., 1993). A mechanism for crossmodal processing has been proposed whereby ventral clastrum mediates communication between different modality-specific perceptual/memory systems (Ettlinger & Wilson, 1990). In humans, neuroimaging studies have provided some evidence for a role of the clastrum in crossmodal processing. A PET study of crossmodal transfer in humans reported that transfer from touch to vision (Hadjikhani and Roland, 1998) is mediated by this area.

Similarly, the strategic location of the insula explains its wide connectivity with the rest of the cortex (see review by Augustine, 1996). Anterior portions of the insula are connected to the frontal lobe, whereas posterior insula is connected to both the parietal and temporal lobes (Tuere et al., 1999). This area also receives gustatory inputs (Baylis et al., 1995: Small et al., 1999) and lesions of this area in humans are known to produce impairments in taste perception (Pritchard et al., 1999). In addition, an fMRI study (Fulbright et al., 1998) and a magnetic source imaging study (Kettenmann et al., 1997) have also implicated this area in olfaction. Right insula activations have also been reported in the context of a PET study of synaesthesia, where passive listening to words evoked colour sensation in synaesthetes (Paulesu et al., 1995).
Subcortical structures - Superior colliculus

Electrophysiological studies report convergence of different sensory streams onto multisensory cells in the superior colliculus (Stein & Meredith, 1993; King & Palmer, 1985; Frens & van Opstal, 1997; Wallace et al., 1996) (Fig. 7), with a suggested role in attending and orienting to different sensory inputs (Sprague & Meikle, 1965; Schneider, 1969; Casagrande et al., 1972; Goodale & Murison, 1975; Stein, 1984; Sparks, 1986; Stein & Meredith, 1990; Meredith et al., 1992). Since this structure receives visual, auditory, and somatosensory inputs (Stein, 1984), from unimodal sensory cortices, convergence of these different types of information might be facilitated. Another relevant aspect of its integrative functions is in processing of temporal information. Multisensory interactions can be maximised by overlapping the periods of peak activity of the unimodal discharge trains. This might be achieved by simultaneous presentations or, in some instances, by non-overlapping sequential presentation (Stein et al., 1996), considering that simultaneous or contiguous stimuli tend to be associated. The
study of processes leading to formation of relations between representations of stimuli or events is a major field in Psychology.

Fig. 7. Coronal section of the human brainstem through the midbrain at the level of the rostral superior colliculus highlighting this latter structure (red) (Afifi and Bergman, 1998).

My Experiments

Chapter 3 addresses the question of what brain structures are engaged during the process of learning word-paired associates, specifically when subjects are instructed to manipulate the encoding process by use of mnemonics. The strategies suggested to the subjects were the use of imagery and semantic associations between the words. I should clarify that this experiment served as an introduction to the study of associative learning and is thus relevant to this thesis. It does, however, differ slightly from the main topic covered in the rest of the
chapters, which tackle questions specifically related to associative learning in the context of multisensory integration.

Chapter 4 is based on a study of audio-visual associative learning and aimed to describe the neuroanatomy underlying this process. Chapter 5 describes an experiment, which followed directly from the one reported in chapter 4. It also addressed audio-visual learning with fMRI but provided more appropriate controls to test for brain structures involved specifically in crossmodal learning, compared to intramodal learning.

Chapter 6 is based on two studies which set out to investigate the question of whether unimodal sensory cortices could be activated by stimuli in modalities other than the one they normally respond to. In one study this question was addressed by using an associative learning paradigm based on pairings between auditory and visual stimuli, and aimed to explore the possibility of visual cortex activating in response to auditory stimuli, which became predictive of a visual stimulus. In the other study, the associations were established between visual and auditory stimuli, and it was hypothesised that auditory cortex might activate in response to visual stimuli, which were predictive of an auditory stimulus.

Chapter 7 is based on an investigation of crossmodal recognition memory. Here, a paradigm was used whereby subjects were exposed separately to visual and auditory study sessions. The material consisted of either pictures or sounds of real objects. Subjects were subsequently tested either in the same modality in which study took place or in the alternate modality. The aim here was to describe the neuroanatomy underlying the process of information transfer from one modality to another when material encoded in one modality was cued for recall in another modality, compared to when study and test occurred in the same modality.
magnetic resonance imaging

Physics

One of the major contributions of technology to medical sciences this century has been magnetic resonance imaging (MRI). The physics underlying MRI is based on nuclear magnetic resonance (NMR). The NMR signal arises from the atomic nucleus. If the number of protons and neutrons in the nucleus is not equal or the total number of protons is odd, an angular momentum or spin is present. The angular momentum allows the nucleus to precess in a magnetic field, and precession facilitates resonance and thus NMR signal. Therefore, only atoms with unpaired protons and/or neutrons can produce an NMR signal. The phenomenon of magnetic resonance is based on the fact that different spin orientations with the magnetic field have different energies, and transitions between spin states can be achieved with radiofrequency electromagnetic waves (Andrew, 1969; Pykett et al., 1982).

The simplest atom in nature is hydrogen, with only one proton in its nucleus. Hydrogen accounts for two thirds of all atoms in the human body and it is also magnetic, i.e. it has spin. Protons with a magnetic dipole moment tend to align in the presence of an external magnetic field. In the absence of such a magnetic field, protons are randomly oriented. Magnetic field strength is measured in Tesla. In brain research MRI magnetic field strength varies between 1 Tesla (T) and 4T.

Spins can have two orientations: parallel and antiparallel to the external magnetic field. The parallel orientation represents the low-energy or ground state, and the antiparallel orientation represents the high-energy or excited state. When placed in a magnetic field, roughly the same number of protons lies in each orientation with a small difference in favour of the ground state. This difference allows spin excitation
and thus causes the high-energy state MR signal, which varies according to the strength of the applied magnetic field, as the population difference is dependent on the field (Stark & Bradley, 1992).

Proportional to the magnetic field strength is the frequency of precession, also known as the resonant or Larmor frequency (Wood & Hardy, 1993). This frequency depends on the type of nucleus. In the case of hydrogen, the precessional frequency is 42.58MHz/T. This frequency increases with the magnetic field affecting the precessing protons. The window of frequencies exploited by MRI lies within the very low frequency, low energy non-ionising radiation, of the radiowave range.

MR signal is detected by radiofrequency (RF) receiver coils when the net magnetisation vector is oriented into the transverse plane. In order to move the magnetisation vector from the longitudinal equilibrium position in the transverse plane (moving spins to the higher-energy level), a second electromagnetic field is applied (an RF pulse) and the spins are brought into phase alignment (Pipe, 1999). The way this is done is by synchronising the applied second magnetic field with the resonant frequency of the precessing protons. The detected signal results from the current induced in the receiver coil by the component of magnetisation vector in the transverse plane. The amplitude (intensity) of the detected RF signal is proportional to the proton density.

The amplitude of the detected RF signal is not constant but decaying (free induction decay). The rate of decay is represented by the relaxation parameters, T1, T2, and T2*, depending on the sequence used. The cause of decay is the gradual loss of coherence (dephasing) between spins as different local magnetic fields are present in different areas of the magnet, resulting in faster spins speeding up and spin protons slowing down. The loss of coherence in turn produces a loss of induced current in the RF receiver coil (Nitz & Reimer, 1999).
In order to bring the spins back into coherence, a 180° RF pulse can be applied. The effect of this second RF pulse is a reversal of direction of precession, regenerating the strong signal in the RF receiver coil. This refocusing of protons by a 180° RF pulse is termed *spin echo*. The rephasing can also be accomplished by rapidly reversing the magnetic gradients, *field or gradient echo* (which, however, does not refocus field inhomogeneities), after which coherence is lost again, leading to an acquisition, reapplication of RF pulses, or reversal of gradients (Pipe, 1999).

Certain effects, however, cannot be corrected by spin echo or gradient echo. An irreversible source of signal loss is due to T2 relaxation effects. These effects, which occur in the transverse plane and are independent of magnetic field strength, are caused by randomly varying intrinsic magnetic fields created by adjacent nuclei in the subject. The interactions between these protons determine the rate of T2 relaxation. T2 may be thought of as the time constant for a first order exponential decay process, or the time required to reduce the transverse magnetisation to 37% of its original value following the RF pulse (Nelson & Runge, 1995; Nitz & Reimer, 1999).

Radiofrequency stimulation essentially adds energy to the system causing the protons to move to a higher excited energy state. However, when this energy dissipates, protons return to the lower energy state, known as T1 relaxation. T1 is the time required for 63% of the longitudinal magnetisation to recover following the RF pulse. Another relevant concept is that of TR or repetition time, marked by the beginning of the first 90° RF pulse through the entire sequence until the delivery of another 90° RF pulse to restart the process (Nelson & Runge, 1995; Nitz & Reimer, 1999).

MR signal intensity, represented by pixel brightness on MRIs, reflects mainly T1 and T2 relaxation times. Other contributing factors are the repetition time (TR) and the echo time (TE, time in ms. from the 90° RF pulse until the echo is received). If
both these times are short, the resulting image emphasises the T1 characteristics of the tissue (T1 weighted), whereas longer TR and TE yield T2 weighted images.

Finally, a critical point in MRI is the signal to (background) noise ratio (SNR). SNR can be improved by signal averaging whereby each acquisition is repetitively observed and summed, and thus increasing the number of observations or excitations. More efficient ways of improving SNR, such as using surface coils, quadrature detection coils, bandwidth reduction, shorter echo times, faster gradient rise times, and optimising system electronics, are available.

**Contrast**

Image contrast in MRI is obtained by virtue of the difference in signal intensity between areas of different composition, arising from an interaction between proton density, T1, T2, magnetic susceptibility, flow and other factors. By rescanning the subject with a different combination of pulse sequence parameters, contrast effects allow MRI tissue characterisation (Nelson & Runge, 1995).

The protons that contribute to the MRI signal are principally found in hydrogen atoms belonging to water and/or fat molecules, also called *mobile protons*. The greater the amount of mobile protons (proton density), the stronger is the signal. However, the strength of the signal also depends on other factors, such as T1 and T2. Among materials with high proton density are fat, cerebrospinal fluid (CSF), blood, and other fluids.

The pulse sequence can only give enough signal intensity when there are enough mobile protons. Certain pulse sequences, such as *proton density* (or spin density) *images* are possible by combining relative long TR and short TE, which results in a decrease of T1 and T2 weightings and thus an increase of the SNR (Lufkin, 1990).
Each material has its own specific T1 and T2 times. Water, for instance, has long T1 and T2 relaxation times, whereas fat protons have relatively short T1 and T2 times. The relevant difference between water and fat is that water has a higher rate of molecular mobile water than fat. For fast relaxation, and thus short T1 time, the Larmor frequency should be tightly correlated to the rate of field fluctuations due to molecular motion. This means that small molecules with high molecular motion, like water, have an inefficiently long T1 relaxation time. However, in the human body, most water is not present in its pure state (bulk water) but bound to proteins and other macromolecules (bound water). Bound water has a slower rate of molecular motion than bulk water, and therefore its relaxation time is optimal (Lufkin, 1990).

On the other hand, T2 relaxation leads to energy exchange between the spins in the excited and ground states, which leads to a loss of coherence of the precessing spins and consequently a reduction of detected transverse magnetisation. The static or low-frequency intrinsic magnetic fields characteristic of large molecules increases the efficiency of T2 relaxation. These intrinsic magnetic fields alter the precession frequency of spins in the local magnetic fields. However, in smaller molecules like water, high precession frequencies average out the intrinsic field to zero and the net magnetic field is determined by the external field, thus maintaining spin phase for longer (Nitz & Reimer, 1999).

Minimising and maximising T1 and T2 can improve the contrast depending on what substance is of interest. For example, lipid-containing molecules, such as fat, have a short T1 (high signal), thus T1 relaxation is directly responsible for the distinction of such tissue in MR. A T1-weighted image results from maximising T1 contrast by shortening the TR time in the pulse sequence or by an inversion pulse, which might lead to a decrease in the signal-to-noise ratio.

On the other hand, T2 relaxation times can affect signal and contrast in such a way as to render certain types of tissue distinguishable. These tissues with a short T2*
relaxation time (low signal) normally contain iron, e.g. deoxyhaemoglobin. Certain pulse sequences, which produce T2-weighted images, result from the increase in image intensity due to T2 relaxation values. In order to maximise the difference in signal intensity based on T2 times, the TE time in the pulse sequence has to be lengthened. If TE is short, the different tissues have similar intensities, and thus little contrast, since not sufficient time has elapsed for differences in T2 to cause dephasing of the spins.

In summary, there is no such thing as an absolute T1- or T2 weighted image. To make a T1-weighted image, an image which reflects the T1 components of the tissue, one has to bear in mind that a short TR maximises T1 contrast and a short TE minimises T2 contrast. To make a T2 weighted image, one has to consider that a long TE maximises T2 contrast and a long TR minimises T1 contrast. The TR and TE values affect the signal-to noise ratio such that this is maximised in the image by a short TE and a long TR (Pipe, 1999).

The degree to which a substance becomes magnetised is called magnetic susceptibility (Lufkin, 1990). Although the atomic nucleus is the basis of MR images, the presence of paired or unpaired electrons creates a magnetic environment, which can also affect the image. Most body tissues have no unpaired electrons but when placed in a magnetic field the tissues show a negative susceptibility (diamagnetic), i.e. a weak magnetic field is induced in the opposite direction to the applied magnetic field. If sufficient unpaired electrons are present, then a strong magnetisation takes place parallel to the applied field, i.e. there is a positive susceptibility (paramagnetic, superparamagnetic, or ferromagnetic) (Forster et al., 1998).

Paramagnetic substances exert their influence on the MR signal by proton-electron dipole-dipole interaction, since the present unpaired electrons produce a greater magnetic dipole due to the greater spin and angular momentum of the electron. This results in an increased efficiency of T1 and T2 relaxation. Paramagnetic
materials have no MR signal of their own but act by altering the proton signal (protons in aqueous solution). In functional magnetic resonance scanning of the human brain, paramagnetic (and diamagnetic) properties of blood are exploited (Forster et al., 1998).

**Functional Magnetic Resonance Imaging (fMRI)**

Neuronal activation produces membrane polarity changes, which result in measurable electric and magnetic changes. Brain activation also increases neuronal metabolism due to energy requirements of membrane repolarisation and neurotransmitter synthesis. These metabolic changes are correlated with changes in blood flow, volume, and oxygenation. This correlation is not yet fully understood. While it is generally accepted that MR signal changes are transduced through neuronally induced haemodynamic changes, the relationship between magnitude, timing, and spatial extent of neuronal activation and haemodynamic changes is not clear. Likewise the relationship between the degree, timing, and spatial extent of induced haemodynamic changes and those of the MR signal changes are not fully understood.

Human brain function can be assessed by detecting the electrical, magnetic, metabolic, and haemodynamic changes associated with neuronal activation. The type of cerebrovascular information of interest here is blood flow. Pauling et al. (1936) observed that the magnetic susceptibility of haemoglobin is sensitive to its oxygen saturation. What causes this susceptibility difference is the diamagnetic oxygen-bound iron in oxyhaemoglobin and the paramagnetic iron in deoxyhaemoglobin.

It is thought that increases in neuronal activity cause local vasodilation and thus an increase in blood flow, which results in an excess of oxygenated haemoglobin and reduces the proportion of paramagnetic deoxyhaemoglobin in the vasculature. This reduction causes a decrease in susceptibility differences around venules, veins
and red blood cells in veins. This decrease leads to an increase in spin coherence and thus an increase in signal in T2* and/or T2-weighted sequences.

The utility of the change of T2* associated with the blood oxygenation change during neuronal activity for imaging purposes became evident from work on cats with induced hypoxia (Ogawa et al., 1990; Turner et al., 1991). Turner et al. found that hypoxia decreased the MR signal from cats’ brains as deoxyhaemoglobin increased (Turner et al., 1991). Moreover, after oxygen was restored, the cat brain’s signal rose above its baseline level due to overcompensation by the vascular system, which brought more oxygen to the blood.

The transverse relaxation rate of blood decreases with an increase in blood oxygenation (Thulborn et al., 1982). Since oxygenation changes lead to small changes in T1 of the whole blood, the primary mechanism of relaxation change is due to intra-voxel dephasing and diffusion of spins through susceptibility-induced gradients (affecting T2, T2* relaxation) and not dipolar interaction (affecting T1 relaxation).

Furthermore, MR signal in the vicinity of vessels and in perfused brain tissue increases if blood oxygenation increases (Ogawa et al., 1990; Turner, 1991). This allows the use of a physiological contrast termed blood oxygenation level dependent (BOLD) contrast, as opposed to a contrast induced by injection of an exogenous substance. Nowadays the most widely employed fMRI technique for the non-invasive mapping of the human brain is high-speed gradient-echo imaging using BOLD contrast (Forster et al., 1998; Howseman & Bowtell, 1999).

BOLD fMRI (which focuses on the T2* change during neuronal activity) and perfusion (which uses the saturation or inversion of incoming blood T1-weighted signal to quantify blood flow) techniques can be used in conjunction in the echo-planar imaging (EPI) MR system (Di Salle et al., 1999). The combination of EPI and fMRI allows for simple collection of T1-weighted images (with EPI) and the EPI
susceptibility sensitivity minimises the T2* imaging. EPI offers a motion-freezing image acquisition speed, which improves fMRI motion artefacts. Also, a long TR for EPI image acquisition helps suppress the MR signal from large vessels. Overall, EPI offers a good signal-to-noise ratio, although its spatial resolution is low.

If one understands TR as the time it takes to acquire a complete volume, be it the whole brain or a section of it, one should aim to use the shortest TR technically possible in order to maximise the temporal resolution. Generally, about 32 slices allow coverage of the whole brain and the TR for such acquisitions lies between 2-3 s. For this duration, one needs not worry for saturation effects, which might be present if the TR was shorter than 2 s, for instance when scanning is restricted to a portion of the brain with a small number of slices. Assuming that one is using multi-slice EPI then the relevant variable is the slice acquisition time (of the order of ~100ms), not the time to acquire an entire volume. The relevant issues are (1) linearity of the evoked response as SOA decreases and (2) ensuring adequate sampling of the haemodynamic response by each slice. Existing evidence suggests that (at least for some cortical areas) linear superposition of haemodynamic responses to successive events starts to break down at <1.5s or so (Friston et al., 1998). So using an SOA of >1.5s would probably be prudent. The second issue regards ensuring each slice of the volume adequately samples the haemodynamic response. This entails avoiding a constant phase relationship between the timing of events and the timing of successive volume acquisitions (because then each slice of the volume would sample only a fixed portion of the haemodynamic response). In practice one can do this by either ensuring that the SOA is a non-integer multiple of the TR (thus making the event onsets progressively 'walk through' the acquired volume) or by adding a 'jitter' to the SOA.

The ISIs for the experiments presented in the following chapters were chosen on the basis of optimizing cognitive performance, i.e. allowing enough presentation duration of items and temporal distance between item presentations to permit appropriate processing of the material. An issue in selecting an adequate ISI is that
too fast a rate of presentation of items could lead to overloading the subjects, such that their responses are not optimal, or if the presentation rate is too slow, subjects might divert their attention away from the task. Another important issue regarding choice of ISI in fMRI experiments is that one has to make sure that scanning takes place during the different phases of the haemodynamic response to items belonging to various conditions. Thus, if the haemodynamic response is on average 5-6 seconds, having an ISI of 5-6 seconds would lead to observations of the same phase of the haemodynamic response. In order to avoid this, the onset of presentations of items can be jittered, such that it is not totally regular, and this does not always coincide with a specific point of the haemodynamic response.

In summary, fMRI has rapidly become popular due to its easy detection of brain activity, safety and non-invasiveness, and its flexibility in terms of scanning duration and the number of scanning sessions that a single subject can undergo. Moreover, BOLD technique combined with EPI allows whole-brain coverage in a very short time (~3 s) (Turner et al., 1998; Ugurbil, 1999). Data collected in recent years serve as evidence of the efficacy of these techniques in elucidating functional maps of the human brain, although it should be noted that resolution is constrained by haemodynamic lag (~5-6 s).

---

data analysis

Various methodological approaches

Analysis of fMRI data can be carried out by use of different software packages. The most widely applied packages are AFNI (Analysis of Functional NeuroImages) developed by Robert Cox at the NIMH in Bethesda, Brain Voyager developed by Rainer Goebel and the Max Planck Society in Germany, FSL (FMRIB Software Library) developed by Steve Smith et al. at Oxford, Freesurfer developed by
Douglas Greve et al. at Harvard University, and SPM (Statistical Parametric Mapping) developed by Karl Friston et al. at the FIL in London.

They are all based on structural transformations of the raw data followed by statistical analysis which allows making inferences from the results, and finally visualisation of functional and structural fMRI datsets. AFNI allows the user to stay close to the data and to view it in many different ways, to make customised analyses, and to add features that can interact with the rest of the package. Brain Voyager is characteristic for its highly optimised 2D and 3D analysis and visualisation routines, with integration of volume and surface rendering for creation of figures and movies, and further tools for brain segmentation, surface reconstruction, cortex inflation and flattening. FSL is self-contained, with easy-to-use graphical user interfaces, with a wide range of structural and functions tools for brain extraction, nonlinear noise reduction, brain segmentation, linear inter- and intramodal registration, unwarping geometrical distortions in EPI images, structural analysis of brain atrophy, and with simple image display utilities. Freesurfer is a tool for cortical reconstruction which allows rendering on surfaces, cortical morphometric analyses, surface-based inter-subject averaging, surface representations, reconstruction of subcortical mass, white/gray and pial surfaces, and topological defects, inflation and spherical morphing, and flattening.

The analysis method chosen for the fMRI data presented in this thesis was SPM. The reason for this choice is simply that the studies were carried out at the FIL, the laboratory where the SPM software package has been developed. Thus, using this data analysis tool was optimal in terms of maximising the support and resources available in this laboratory. Thus, what follows is a detailed description of SPM.

**Univariate versus Multivariate approaches**

Functional magnetic resonance imaging (fMRI) works under the assumption that haemodynamic response is a correlate of neural activity. Evoked haemodynamic
responses impose certain spatio-temporal limits (2–5mm and ~5-6 s). FMRI has a spatio-temporal scale of 1–3mm (of brain surface) and 1sec. Thus the effective resolution of fMRI is constrained by physiological limits. FMRI can measure brain state–related and event–related responses. If presentation of stimuli is organised in epochs or continuous task performance, then the underlying signal is interpreted as a brain–state dependent measure. By contrast event–related responses can be measured from presentation of single stimuli.

At the theoretical level there are two principles which govern brain organisation: functional integration (multivariate) and functional segregation (univariate). Integration refers to interactions between functionally segregated areas. Segregation suggests that cortical areas, which are specialised for some computation, are anatomically segregated. This concept differs from that of functional localisation, which implies that a function can be localised to a cortical area. The reason why these two distinctions are relevant is because they constitute the basis of different analysis models for imaging data, and the interpretation of functional imaging studies relies on the implementation of such models.

Functional integration approaches are based on correlated physiological dynamics in different parts of the brain (multivariate), and deal with distributed brain systems. Examples of multivariate approaches are: eigenimage analysis, multidimensional scaling, multivariate analysis of covariance (ManCova), and canonical variates analysis (CVA). On the other hand, functional segregation approaches are based on detection of focal differences. These univariate approaches find their main exponent in statistical parametric mapping (SPM). SPM addresses the significance of regionally specific effects by treating each voxel separately (univariate). By performing voxel–based statistical analyses in parallel, it creates an image of a statistic. It is precisely these image processes, with voxel values distributed according to a known probability density function \( t \), which result in statistical parametric maps.
Statistical parametric mapping originates from two existing ideas: change
distribution analysis, i.e. voxel–based assessment of neurophysiological changes
developed for PET studies (Fox & Mintun, 1989), and significance probability
mapping, adapted from the analysis of multichannel EEG data involving the
construction of interpolated pseudomaps of a statistical parameter.

**Spatial transformations**

The stages of analysis of fMRI data (echo–planar imaging) are: realignment,
spatial normalisation, spatial smoothing, voxel–based statistical analysis (general
linear model), and statistical inference (theory of Gaussian fields).

The first step is **realignment**, where movement–related variance components,
which can confound the analysis, are removed. This step is performed by
determining a rigid body transformation for each of the images, registering all
functional images to the first in the series. Realignment results in an optimisation of
the parameters by optimising the residual sum of squares, and is a within-modality
operation (only functional image series are used).

Realignment might prove a difficult operation since the movement in earlier scans
can affect the signal in subsequent scans. Head movements are estimated and
then used to realign the images and adjust for movement–related components that
persist after realignment. This adjustment consists of a moving average–auto–
regression model of spin excitation history effects. Sometimes it is convenient to
register a structural image to the functional scans. This is another rigid body
registration but since the structural image is acquired in a different modality to the
functional images, the registration is not a simple minimisation of the residual sum
of squares. These between-modality registrations are performed by partitioning the
images into grey and white matter and then simultaneously registering the
partitions together. This step is usually termed **coregistration**.
In order to perform voxel-based analysis, the data need to be in the same anatomical space, i.e. the data from different subjects must derive from homologous brain regions. In the normalisation step, the images are subject to specific spatial transformations such as nonlinear warping, where the optimum parameter affine transformation is determined in order to match a template which conforms to a standard coordinate system, that of the Montreal Neurological Institute (MNI) (Evans et al., 1992, 1993). Also, the template image is in the same modality as the images to be registered so the optimisation consists of a minimisation of the residual sum of squares. The parameter transformation corrects for the variation in position and size of the image (more subtle effects are corrected by a nonlinear registration).

**Smoothing** aims to increase the signal to noise ratio and involves convolution of the data with a smoothing kernel. In fMRI the noise can be regarded as fairly independent for each voxel and thus has high spatial frequency components. Also, if the convolution uses a Gaussian kernel, the data will conform more closely to a Gaussian field model, which is relevant if statistical inference employs theory of Gaussian fields to assign p values to regionally specific effects. Finally, smoothing is important in intersubject averaging to ensure that haemodynamic changes between subjects are assessed spatially where functional anatomy homologies can be observed.

I generally applied a smoothing kernel of 6mm for each subject. For the first data set I applied a 10mm smoothing filter constituted my first analysis. These parameters were commonly used at the time (see Fletcher et al., 1998a, 1998b; Büchel et al., 1998b). For subsequent datasets and also as my sample sizes increased, justify using this filter was no longer justifiable, and the smoothing filter became more conservative.

After smoothing, univariate tests are performed at the voxel level by means of the general linear model. The resulting test statistics constitute a statistical parametric

70
map. The data can now be divided into neurophysiological responses due to components of interest, confounds, and an error term. This partition takes place when the parameter estimates associated with the design matrix are estimated using the general linear model. The contribution of each parameter estimate is assessed in terms of a $F$ or $t$ value for each voxel, $\text{SPM}(F)$ or $\text{SPM}(Z)$. $\text{SPM}(F)$ is used to make inferences about several effects simultaneously, whereas $\text{SPM}(Z)$ informs about single effects. Temporal smoothing also aids to maximise signal components.

Finally, **statistical inferences** are made on the basis of the SPM and the responses observed in terms of fitted responses or parameter estimates. When there is an *a priori* anatomically constrained hypothesis about an effect in a specific brain area, the inference can be based on the statistical value without correction, the $Z$ value in that region in the $\text{SPM}(Z)$ is enough to test the hypothesis. However, if anatomical sites have not been predicted to activate, then correction for the multiple dependent comparisons performed is required. Corrections are made by application of distributional approximations from the theory of Gaussian fields. Corrections can also be implemented for predicted regions using small volume correction (Worsley, 1995). This latter approach is usually implemented when testing whether a specific activation falls on a region demarcated either by anatomical characterisation or previous findings of functional localisation.

Thus, in this thesis, all corrected data are reported regardless of whether they were expected or not. These criteria are clarified in the general methods chapter (pag 65). A small volume correction is used in the study reported in chapter 5 to delineate the overlap of the hippocampal activation previously found in the study described in chapter 4. This was important in order to highlight the relevance of this region in similar crossmodal learning paradigms. At times, activations were found to be significant at the cluster level, coinciding with *a priori* hypothesised regions, which is why they were reported. That they reached correction at the cluster level was reported for completeness but not as a means to justify the significance of the
activations, such as in the case of voxel-level correction. Finally, it should be clarified that in reporting uncorrected data, arbitrary thresholds (e.g. all clusters with voxels size greater than 10 voxels) were used in order to cut the list at some sensible point. In one study (chapter 7) this threshold was 10 voxels, whereas in another (chapter 4) it was 40 voxels. These happened to be the thresholds that corresponded with the observation of expected activations. The threshold is arbitrary (as with all thresholds) and would be sensibly chosen to exclude very small blobs that are less than the resolution of the imaging device. This would be dictated by the smoothness and would translate as between one and a half and twice the full width at half maximum. In this sense, it was also the case that the smoothness used in the study reported in chapter 4 was greater than that used in the study reported in chapter 7.

The general linear model

The general linear model is an equation, which contains variables for the observed response in a linear combination of explanatory variables and an error term. The general linear model can come in various forms: analysis of covariance, t-test, or multiple regression analysis. It basically compares what is observed to what is expected by expressing the observations or response variables as a linear combination of expected components or explanatory variables, and a residual error. Thus, for a response variable \( X_{ij} \), such as cerebral blood flow (rCBF) at a given voxel \( j=1, J \), the general linear model corresponds to the following equation:

\[
Y_{ij} = X_{i1}\beta_{1j} + X_{i2}\beta_{2j} + \ldots X_{ik}\beta_{kj} + \epsilon_{ij}
\]

where \( i = 1,\ldots, I \) refers to the specific observation or scan. Errors \( \epsilon_{ij} \) are assumed to be independent and normally distributed, which in activation studies implies an equal error variance across conditions and subjects. \( \beta_{kj} \) refers to \( K \) unknown parameters for each voxel \( j \) representing the relative contribution of each explanatory variable. The coefficients \( X_{ik} \) are explanatory variables referring to the
conditions under which the observation was made. These coefficients can be covariates or indicator-type/dummy variables. The design matrix can contain both types of coefficients, where each column of the design matrix has an unknown parameter in the $\beta_j$ vectors, some of which will be of interest and some will be of no interest or pertain to confounding effects.

By using the general linear model, one can express the relative contributions to the BOLD response of the different variables. These contributions are expressed as the parameter estimates of the model. The actual numerical values of the parameter estimates are arbitrary, i.e. they do not represent a quantitative measure of BOLD signal, but rather a relative difference in BOLD signal between the different covariates. The data can be further presented in graphical form by plotting the parameter estimates corresponding to various conditions of interest and observing the relative contributions to the observed effect.
The design matrix

The theoretical frameworks encompassed by SPM are the general linear model and the theory of Gaussian fields (Friston et al., 1995c). The experimental design and the model used to test for specific neurophysiological responses are embodied in a mathematical structure called the design matrix. The design matrix is partitioned into effects of interest (activation) and effects of no interest (e.g. global activity or other confounds). These effects can reflect factor levels, such as the presence of a specific cognitive component, or they can be continuous functions, such as reaction times. The design matrix is constituted by a series of columns, each of which represents one effect. The first few columns of the design matrix represent the conditions from which the response of interest is to be extracted,
followed by some terms, which aim to remove or model low frequency artefactual variations in signal. The final column is generally an index of whole brain activity.

\[
y = X \beta + \epsilon
\]

Fig. 2. Diagram showing the different components of the design matrix (courtesy of Stefan Kiebel, 2001). In this case, from a time series, two conditions are accounted for by \(X\), a variable of interest and a session effect, thus two beta values (parameter estimates) indicate the contribution of each condition to the observed effect.

The contribution from each effect in the design matrix is estimated using the general linear model and standard least-squares tests. These contributions are termed parameter estimates. The differences between parameter estimates reflect regionally specific effects, such as a task-related activation, by use of linear compounds and contrasts. The significance of a contrast is tested with a Student distribution statistic under the null hypothesis. For each contrast (differences between parameter estimates) a statistic is computed for each voxel to yield a SPM\(t\). When an SPM\(t\) is transformed into a Gaussian field it becomes a SPM\(Z\). Statistical inferences can be made on local aspects of the SPM\(Z\) by specifying a threshold of distributional approximations from the theory of Gaussian fields. Such distributions characterise the sites of activation in terms of maximal value and/or spatial extent.
Statistical inferences within SPM are based on the probability of obtaining corrected p values at different levels. A specific formulation assigns different types of corrected p values pertaining to: the number of clusters (c), the size of clusters (k), and the height of voxels within a cluster (u).

Experimental designs

By specifying different contrasts in the design matrix, various effects can be tested. There are three major types of design: parametric (dimensional), factorial (interaction), and categorical (subtraction–conjunction) designs. In parametric designs, regional physiology is modelled in a nonlinear and systematic fashion testing for variations in the degree of cognitive and sensorimotor processing (Grafton et al., 1992; Price et al., 1992). In contrast, factorial designs combine various factors in the same experiment. The effect of one factor on the effect of another factor can be assessed by an interaction term. Interactions represent a change in a change, which is a useful approach when two or more factors are combined and are possibly influencing each other. Factorial designs can also assess time by condition interactions (Friston et al., 1992a) observed in processes of physiological adaption and plasticity (Gilbert & Thach, 1977), or in modulatory processes such as psychopharmacological activations (Friston et al., 1992b). Finally, in the subtraction–conjunction approach, the difference between two tasks is formulated as a separable cognitive or sensorimotor component and the corresponding functionally specialised area can be identified by specific differences in brain activity. Subtraction tests a hypothesis related to the activation in one task relative to another. Thus, several hypotheses about the joint significance of activations in task pairs can be tested. If there is one single component resulting from the task pair-differences, then that region of activation can be associated with that specific component. Conjunctions, on the other hand, work through a combination of a series of subtractions. The main differences between subtractive and parametric approaches is that cognitive processes are not treated as categorical invariants (subtractive) but as a dimension
that can be expressed to different extents. Finally, factorial designs are a useful tool in the assessment of the effect of one manipulation over the effects of another.

**Fixed versus Random Effects Model**

The manner in which data are obtained always qualifies any derived inferences. Here, examples are used to describe two types of data modelling: fixed and random effects models. In the case of fixed effects, consider a home gardener carrying out a small experiment with 24 tomato plants, comprising 6 plants of 4 varieties that the gardener is particularly interested in. Comparison of the four varieties is now to be made in the 12' x 8' garden space available. Each plant is allocated randomly to one of the 2' x 2' squares. If $y_{ij}$ is the yield of fruit from plant $j$ of variety $i$, a possible model for $y_{ij}$ would be:

(1) \[ E(y_{ij}) = \mu_i \]

where $E$ represents expectation and $\mu_i$ is the expected yield from a plant of variety $i$. This can be transformed into:

(2) \[ E(y_{ij}) = \mu + a_i \]

where $\mu$ is a general mean and $a_i$ is the effect on yield of tomatoes due to the plant being variety $i$.

In this modelling of the expected value of $y_{ij}$, each $\mu_i$ is considered as a fixed unknown constant, the magnitudes of which we wish to estimate. In doing this, $\mu$s correspond to the four different varieties that the gardener is interested in. They are four very specific varieties of interest, and in using them the gardener has no thought for any other varieties. This is the concept of fixed effects. Attention is fixed upon just the varieties in the experiment, upon these and no others, and so the effects are called *fixed*. Because all the effects in (2) are fixed effects, the model is called fixed effects model.
This example describes a sampling process pertinent to this fixed effects model. The data are envisaged as being one possible set of data involving these same tomato varieties that could be derived from repetitions of the experiment, repetitions for each of which a different sample of 6 plants of each variety would be used. The most important feature of fixed effects is that they are deemed to be constants representing the effects on the response variable of the different levels of the factor concerned, in this case the variety of tomatoes. These varieties are the levels of the factor of particular interest, chosen for the interest in those varieties in the experiment.

In contrast, an example of random effects models is offered to illustrate the basis of this different approach of data analysis. Suppose a new form of injectable insulin is being tested using 15 different clinics. It is not unreasonable to think that those clinics as a randomly chosen sample of clinics from a population of clinics. If clinic $i$ has $n_i$ patients in the trial and the measured response of patient $j$ in clinic $i$ is $y_{ij}$, then a possible model would be:

$$E(y_{ij}) = \mu + \alpha_i$$ for $i=1,...,n_i$.

Although equation (3) is algebraically the same as (2) from the previous example, some assumptions underlying it are different. In the first example each $\alpha_i$ is a fixed effect, $i$ being pre-decided as an effect of interest. However, in (3) $i$ is the effect on blood-sugar level of the observed patient having been injected in clinic $i$; and clinic $i$ is just one clinic, from among the randomly chosen clinics that happened to be numbered $i$ in the clinical trial. Since the clinics have been chosen randomly with the object of treating them as a representation of the population of all clinics, and from which inferences can be made about that population, the one labelled $i$ is of no particular interest of itself to the trial; it is of no interest solely as being one of the 15 clinics randomly chosen from a larger population of clinics. This is characteristic of random effects: they can be used as the basis for making inferences about populations.
In the context of the studies presented in the ensuing experimental chapters, neuroimaging data are aimed to represent effects that pertain to the general population. Thus, most of the data analysis is based on random effects model. The first study in chapter 4, which was the one performed first chronologically, includes a small sample (5 subjects) and was analysed using a fixed effects model. Thus, the inferences drawn pertain to this specific subset of the population, and not to the general population. The reason for this approach in this instance was a simple methodological limitation, in that random effects modelling of data had not yet been implemented for SPM.

The standard sample size for a fixed effects analysis was 6 subjects and 12 for a random effects analysis. For the first study (chronologically), reported in chapter 4, I initially scanned 6 subjects but had problems with one participant, and excluded him from the group. For the rest of the studies, random effect analyses were carried out and my sample size was always at least 12, except for the study reported in chapter 3, where I excluded one subject whose behavioural data were significantly poorer from the rest. In future, I think the sample size should not be chosen based on currently acceptable thresholds but rather on levels of significance. Thus, a subtle effect might still not be significant with 12 subjects but might be observable with a larger sample.

**Main versus Time-dependent Effects**

Two main types of effects are used to describe neural responses. Main effects constitute an average of neural response over a period of time (e.g. a session) for a given condition relative to a baseline condition, a control condition, or another condition of interest. In this respect, main effects give an estimation of brain regions involved in a specific task regardless of time effects on activations of specific areas. A complementary approach is that of time-by-condition interactions, which portray neural responses evoked by a specific condition modulated by time. This method of data analysis allows investigation of increases and decreases of
brain activation over time for a given condition relative to a baseline or for various conditions relative to one another.

In the study of learning processes, time is considered a key factor. Brain activations are expected to change with time as learning takes place. From a neural perspective, learning is assumed to be based on patterns of brain activity, which differ before and after new material is encoded. Neuroimaging affords us the possibility to trace these changes in terms of BOLD response variations. Several neuroimaging studies have already reported the use of such data manipulations for the purpose of studying change in brain activation (Büchel et al., 1998a; Büchel et al., 1999a; Fletcher et al., 1999; Raichle et al., 1994), although it remains a fairly novel analytical approach. It is precisely this type of data analysis that constitutes the bulk of the fMRI results reported in the experimental chapters that follow.

Changes in BOLD response can be either increases or decreases. BOLD responses have shown to reflect synaptic activity (Logothetis et al., 2001; Rees et al., 2001*). Increases in BOLD response could thus be enhanced excitation or inhibition. Decreases in BOLD responses remain controversial. Decreases could show reduction of response, such as by familiarity or habituation. Alternatively, decreases could also be a consequence of suppression, such as in the case of crossmodal suppression (Laurienti et al., 2002).

On a related note, in order to study changes in BOLD signal over time as a key feature of learning processes, it seemed pertinent to split the repetitions of material into temporal partitions or sessions, so as to compare the differences in activations observed at different points of the experiment, normally during an initial, intermediate, and final or more practiced phase. In particular, the experiment described in chapter 5 required partitioning of presentations into sessions since the material in each session varied, having audiovisual pairs in one session, auditory-auditory pairs in another session, and visual-visual pairs in another. The aim of such partitions was to avoid confusing subjects with too many items paired in different arrangements. However, the contrasts performed subsequently included
all conditions, regardless of the session in which they presented. In the case of the experiment presented in chapter 7, for instance, a different strategy was adopted, whereby the experiment was conceptually divided into two sessions, one corresponding to the training phase and the other to the test, but scanning being continuous with two different screen displays signaling the beginning of the training and the test phases.

And finally, when plotting results graphically, I have referred to values on the y-axis as either signal change or effect size. These values were either the estimated parameters themselves or a linear combination of them (as specified by a contrast). Therefore they indicate both sign and size of the BOLD signal change as fitted by the regressors in the design matrix. Provided evoked hemodynamics are appropriately modelled by the regressors in the design matrix, then the beta values are a direct indication of the strength of the haemodynamic response in each voxel, or in other words of the underlying signal change. Since these values represent a relative measure, units are arbitrary (a.u.) and not specific.
CHAPTER 3: A Neuroanatomy of Mnemonics

introduction

Imagery allows us to internally represent the world in the absence of sensory input and is useful in various cognitive tasks, including learning. Paivio (1969) has demonstrated that concrete words, which are more imaginable than abstract words, are more easily learnt. Thus, imagery can be used to enhance learning, and the interactive imagery technique, which instructs subjects to imagine two objects interacting, has subsequently been shown to enhance memory as measured in free recall (Boir, 1970; Schnorr and Atkinson, 1970). Several theories have attempted to explain the effects of imagery on learning. The dual coding hypothesis (Paivio, 1986) asserts that learning paired-associates involves the interaction of two interconnected, but functionally independent, systems: a non-verbal imagery system and a verbal system. This psychological theory can account for findings that concrete words, due to their capacity to induce activity in both verbal and nonverbal processing systems, are more readily learnt than abstract words, only available to a verbal code (Paivio, 1971).

In neuropsychological studies, imagery has been used as a tool to aid learning in memory impaired patients. Kovner et al. (1983) developed a learning technique for chronic amnesics to link items through ridiculous-image stories (RiS), thus explicitly exploiting the use of imagery. The method of interactive imagery has also been shown to improve performance in free recall of concrete and abstract nouns in head injury patients (Richardson and Barry, 1985; Twum and Parente, 1994). Additionally, neuroanatomical evidence from brain lesioned patients suggests that when the temporo-occipital region involved in visual memory functions is damaged,
patients fail to benefit from imagery in memorising concrete words (Goldenberg, 1989).

A contrasting theory to the dual-coding hypothesis has argued that retaining information as pictures would exceed human storage and retrieval capabilities. Consequently, the conceptual-propositional hypothesis (Anderson and Boir, 1973) proposed that information is stored in an abstract propositional format. This contrasting hypothesis carries an implication that words are stored in a verbal code, where enhanced richness in the set of predicates which bind concepts together, explains the advantage for concrete, relative to abstract, words during learning.

Neuropsychologically, a distinction between a verbal and a nonverbal visual semantic system has also been made (Warrington, 1975; Schwartz et al., 1979). This is illustrated by the cases of AB and EM (Warrington, 1975) and WLP (Schwartz et al., 1979), who were able to visually identify objects correctly but experienced difficulty retrieving the meaning of words relating to the same object. Further evidence of separate semantic systems comes from modality-specific aphasias (Spreen et al., 1966; Lhermitte & Beauvois, 1973; Assal & Regli, 1980; Beauvois, 1982; Poeck, 1984; Gil et al., 1985; Denes & Semenza, 1975; Beauvois et al., 1978). It has been proposed that impairment in communication between separable verbal, visual non-verbal (and tactile and auditory nonverbal) systems might be the underlying cause of these types of aphasia (Beauvois, 1982). However, regarding analysis of abstract properties, a distinction between a verbal and a nonverbal semantic system might be more difficult to sustain. In fact, it has been suggested that a purely verbal semantic system is insufficient for adequate comprehension, and that it acts as an input to other semantic systems (Johnson-Laird, 1983). Thus, although verbal and nonverbal (e.g. visual) semantic systems might be to some degree separable, it seems implausible that these are totally autonomous from each other, and they might be more convincingly viewed as coupled subsystems (Shallice, 1988).
Recently, neuroimaging techniques have provided information concerning neuroanatomical substrates of imagery and imagery-based learning and memory. Several fMRI studies have reported involvement of parahippocampal gyrus during encoding where the material consisted of concrete items (Wagner et al., 1998b; Breir et al., 1998; Krause et al., 1999a). Fletcher et al. (1995a, 1996) reported medial parietal cortex activation during retrieval of imageable, relative to non-imageable, word paired-associates. Generally, the medial parietal region has been implicated in visual memory retrieval (Tulving et al., 1994b; Shallice et al., 1994; Fletcher et al., 1995a; Buckner et al., 1996a) with the suggestion that projections from inferior temporal cortex (Harries & Perrett, 1991) facilitate access to imageable representations from visual memory stores.

In addition to the dimension of imageability, semantic organisation of material has a facilitatory effect on learning (i.e. as assessed at subsequent retrieval). For instance, during word paired associate learning, pairs containing high associates (e.g. table-chair) are more likely to be recalled than low associates (Jenkins & Russell, 1952; Deese, 1959). Furthermore, organisation involving category-exemplars (categorisation) leads to better recall than randomly organised word lists, where the order of retrieval tends to be according to category clusters (Bousfield, 1953; Tulving & Pearlstone, 1966). In a study using words arranged in hierarchical patterns, such that each word converged into a meaningful hierarchy (e.g. platinum converged, along with silver and gold, on rare; which in turn converged, along with common and alloys, on metals; which in turn converged, along with stones, on minerals), it was reported that retrieval of semantically organised words was significantly higher, relative to words which had no underlying organisation (Boir et al., 1969). It is also proposed that when material to be encoded is randomly presented, subjects impose their own structure, a phenomenon described as "effort after meaning" (Bartlett, 1932). Thus, when words are studied with no pre-specified organisation, retrieval occurs in a consistent order, implying an active imposition of organisational schemes (Tulving, 1962, 1964).
The idea of semantic association necessarily implies degrees of linkage between items. The associationist concept of *semantic distance* was introduced in the context of the spreading activation theory (Collins & Loftus, 1975), which proposes that highly related concepts are located close together and the shorter the distance between them the more easily excitation flows from one node to the next. Psychological evidence that bears on the validity of the concept of semantic distance includes findings that certain words (e.g. *butter*) are more rapidly identified when followed by closely associated words (e.g. *bread*) than when followed by distally or non-associated words (e.g. *doctor*) (Meyer & Schvaneveldt, 1971).

From a neuroanatomical perspective, it is likely that different brain systems are activated when learning paired-associates having a semantic or meaning-based association relative to unrelated items. The former process is likely to involve automatic retrieval of pre-existing memories, whereas the latter is likely to engage active search or generation of new linkages. It follows from this that retrieval processes may in some instance occur in parallel with encoding processes.

A region of ventrolateral prefrontal cortex (VLPFC) is proposed to mediate ‘deep’ processing of items, by means of semantic tasks (Kapur et al., 1994a, 1994b). More recent fMRI studies reinforce the role of left inferior frontal cortex in memory encoding (Fletcher et al., 1998a; Menon et al., 2000), more specifically during semantic encoding (Wagner et al., 1998a) and during verb generation in relation to concrete nouns (Thompson-Schill et al., 1998). Verb generation in language necessarily involves creating meaningful associations involving some kind of action. Thus, VLPFC might participate in semantic organisation through various processes, including active retrieval of semantic or deeply encoded knowledge, verbally based search for meaningful associations, and selection between competing items of information.

In this study I address the modulatory role of imagery and semantic associations on patterns of brain activation during learning. The question formulated was
whether patterns of activation during paired-associate learning can be differentiated according to attributes of the studied material. An additional consideration was whether neuroimaging data might yield evidence to support dual-coding or verbal-propositional theories of encoding. A priori hypotheses for this study are based on previous findings of brain areas involved in processing concrete/abstract and semantically associated/unassociated items, reviewed in this introduction.

---

**methods**

**Subjects**

Eleven healthy right-handed (Edinburgh handedness test) volunteer subjects (5 females) whose ages ranged from 23 to 36 years participated in the study. All were free of neurological or psychiatric illness. All subjects provided informed consent. This study was approved by the joint Ethics Committee of the National Hospital for Neurology and Neurosurgery and the Institute of Neurology, London.

**Experimental materials**

Subjects were presented with word paired-associates, all of which were nouns. The experimental design comprised two factors: imageability and semantic associatedness. Thus items in a pair were either imageable or non-imageable, and semantically-related or semantically-unrelated, resulting in four different conditions, namely (1) imageable-related e.g. airport-plane, (2) imageable-unrelated e.g. mouse-cabbage, (3) nonimageable-related e.g. future-past, and (4) nonimageable-unrelated e.g. choice-grace. A fifth baseline control condition involved the use of pairs of letter strings (nonwords).
All word pairs were created using the Birkbeck word association norms (Moss & Older, 1996). Words were counterbalanced for number of letters in the word pair and the average frequency ratings were: 81.425 (Kucera-Francis Written Frequency scale, 1967), 452.475 (Thorndike-Lorge Written Frequency scale, 1944), and 9.1 (Brown Verbal Frequency scale, 1984). The average semantic relatedness between word pairs was 31.335, according to the scale of Birbeck word association norms. The average imageability rating was 325.625, as derived from merging the Paivio, Colorado, and Gilhooly-Logie norms (Coltheart, 1981), and the average word pair length was 6.45 letters. The order of conditions was randomised, counter-balanced for word length, imageability ratings, frequency and familiarity.

Cognitive task
The task consisted of encoding word-pairs by making meaningful associations between the items. In the case of semantically related words, the association was obvious (e.g. elephant-trunk), whereas in the case of semantically-unrelated word-pairs (e.g. carnation-sword), subjects were instructed that in the absence of an obvious link they were to establish an association, either through imagery or by links in memory, that would aid subsequent recall of these items at test. At test subjects were given a cued-recall task whereby a word was presented and its matched pair had to be retrieved through recall.

Experimental Procedure
Subjects underwent three consecutive study sessions while being scanned where the paired-associates were presented visually in an event-related design using a pseudorandomised presentation pattern. Stimuli were presented in white against a black background and centred on a screen located approximately 20 cm. away from the eyes, subtending a visual angle of 7 degrees.
There were 20 word-pairs for each condition repeated three times during the whole experiment, i.e. once per study session. The presentation duration for each pair was 1.3s, with five stimulus types being randomly presented in blocks of five. The inter-stimulus interval, which was randomly jittered ($\pm 0.5s$ and $\pm 1s$) to avoid expectation effects, was 9s on average. The duration of the entire experiment was approximately 60 min.

**Magnetic Resonance Procedures**

The data were acquired from a 2-Tesla Magnetom-VISION whole-body MRI system (Siemens, Erlangen) equipped with a head volume coil. Multislice T2*-weighted fMRI images were obtained with a gradient echo planar sequence and axial slice orientation (echo time TE=40s, repetition time TR=4.1s, 64x64x48 voxels). Data were acquired in three sessions. A total of 675 contiguous volume images were acquired per experimental subject. Each echoplanar image comprised 48 1.8 mm. slices taken every 3 mm, positioned to cover the whole of the cerebrum. Each session comprised 225 volume images.

The functional scans were acquired in an event-related fashion. The scanner was synchronised with the jittered onset of events, and the ratio of interscan to interstimulus interval ensured that voxels were sampled at different phases relative to stimulus onset (a total of 2.25 scans were taken per event).

Structural images were obtained in the sagittal plane, yielding T1-weighted images with 1x1x1.5 mm. voxel size (matrix size: 256x256x108). Structural images were acquired before the start of the experiment, followed by one or two scout scans and five dummy scans, before the first functional scan, to allow for T2 equilibration effects.
Data analysis

The data were analysed using the general linear model, in this instance linear regression on condition-specific waveforms convolved with a haemodynamic response function (Friston et al., 1995a) as implemented in SPM98 and SPM97devel. More specifically, the time series were realigned and resliced using sinc interpolation, adjusting for residual motion-related changes, spatially normalised (Friston et al., 1995b; Friston et al. 1996) to a standard EPI template (Evans et al., 1992, 1993; Collins et al., 1994), and smoothed spatially with an 8mm. FWHM isotropic Gaussian kernel and temporally with a 4s Gaussian kernel, and a high-pass filter of 56s.

The analysis was based on data gathered for all the learning events, without exclusion of events not leading to subsequent memory. A group analysis was performed involving a first-level analysis, wherein a single image was created for each condition and for each subject. The regressors included in the design matrix were main and time-dependent effects for each condition, each subject and each session. Subsequently, a second-level analysis group comparisons were drawn following a random-effects model. Reported activations either survive voxel-based and/or cluster-based correction for multiple comparisons (p<0.05) or were a priori hypothesised on the basis of previous findings. In the latter case the threshold for reporting data is p<0.001 uncorrected, and p<0.005 uncorrected for one exception. The maxima of these areas on a T1 template brain were labelled following the nomenclature of Talairach and Tournoux (1988), Duvernoy (1999), and Brodmann (1909).
results

Behavioural data

Subjects were tested outside the scanner following three presentations of the material at study, which took place during scanning. The test consisted of a cued-recall task where a word was presented and its paired match had to be retrieved. The mean percentage of correct responses for the group was 91%, which exceeds chance levels (p<0.05). The mean performance on each condition is illustrated graphically (see figure 1).

Fig.1. Mean behavioural performance shows the percentage scores from the cued retrieval task, which followed the encoding sessions inside the MRI scanner.

An analysis of two-factor (imageability and semantic relatedness) with replication analysis of variance (ANOVA) revealed no significant difference (p>0.05) in performance among study conditions. However, two-sample t-tests performed between the nonimageable-semantically unrelated condition and all other conditions revealed a significantly lower performance scores in the former condition (p<0.05). Paired-associates belonging to this condition were predicted to be the hardest to encode since the items in the pair were neither imageable nor semantically related. Thus, performance in this condition might have been dependent on successfully finding memorable self-generated associations between these words.
Neuroimaging data

The interest in this study was in learning and not in retrieval, thus it seemed pertinent to include all learning events. Had the interest been in successful retrieval, learning events leading to unsuccessful retrieval should have been excluded. Also subtracting events would have unnecessarily reduced the statistical power, which would have preented a bias for the most difficult condition (non-imageable, non-associated pairs). Thus, by excluding these events, a confound of level of difficulty/effort would have been introduced, and testing for this factor was not the aim of the study.

However, it is pertinent to control for performance differences because these directly measurable changes in behaviour certify that (1) learning has taken place, and (2) that the activations I am reporting are most likely correlated with the neural processing underlying the change in behaviour. Controlling for performance differences can also inform of subjects, who did not perform the task properly by giving significantly different scores to the rest of the group.

Functional MRI data were examined with respect to two main dimensions: imageability and semantic relatedness. In the first instance, main effects were computed for the effect of imageable versus nonimageable material at encoding (and nonimageable versus imageable) regardless of semantic relatedness, and the effect of semantic relatedness versus unrelatedness (and vice versa) regardless of imageability. The fifth condition, i.e. meaningless letterstring (nonwords), was used as a control on preliminary contrasts to test the general validity of the data. For this purpose, several comparisons between pairs of meaningful letterstrings (words, i.e. the other conditions) and pairs of meaningless letterstrings (nonwords, i.e. control condition) revealed activation of well known language areas.

Simple effects were also assessed. Apart from these main comparisons, interactions between the factors of imageability and semantic relatedness were
also examined, although no effects were observed that survived a threshold for corrected significance of $p<0.05$. This was also the case for the time by condition interactions, which were examined for differential time-dependent changes in activation as a function of learning.

Results reported here focus on main effects surviving correction ($p<0.05$) at the voxel-level (indicated by *) and/or at the cluster-level (indicated by †). All other results involving a priori predicted regions of activation were thresholded at $p<0.001$, uncorrected, with one exception where $p<0.005$ uncorrected was used for reporting simple effects. Simple effects (comparisons between single conditions) are generally reported using the same significance thresholds, except for a case where the areas found using a $p<0.005$ had been observed in the main effects (generated from comparing pairs of conditions).

a) Imageability versus Non-imageability
When comparing the imageable with the nonimageable conditions, independent of semantic relatedness, main effects were found in bilateral parahippocampal gyrus (BA36), right posterior cingulate gyrus (BA23/30/31), right inferior frontal gyrus (BA46), and bilateral angular gyrus (BA19) (see fig. 2). Note that many of these effects are highly significant, surviving correction at $p<0.05$ (see table 1). Examination of parameter estimates shows a difference in the relative contribution from different conditions to the main effect in the parahippocampal region (fig 2.1 b2). From the graph a difference in activation between the two imageable conditions is obvious and was subsequently confirmed by direct comparisons of simple main effects of imageable nonsemantic versus imageable semantic pairs which revealed greater left parahippocampal activation ($p<0.001$, uncorrected) associated with the nonsemantic imageable condition.
Table 1. Brain regions revealed from the comparison of imageable versus non-imageable word paired-associates

<table>
<thead>
<tr>
<th>Brain area</th>
<th>Coordinates (x y z)</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>R posterior cingulate gyrus BA23/30/31</td>
<td>12 -50 20</td>
<td>6.08*†</td>
</tr>
<tr>
<td>Bilateral parahippocampal gyrus BA36</td>
<td>-30 -28 -18</td>
<td>5.94*†</td>
</tr>
<tr>
<td></td>
<td>34 -20 -24</td>
<td>4.96*</td>
</tr>
<tr>
<td>L angular gyrus BA19</td>
<td>-36 -78 36</td>
<td>4.97*</td>
</tr>
<tr>
<td>R angular gyrus BA19</td>
<td>42 -70 40</td>
<td>4.55†</td>
</tr>
<tr>
<td>R inferior frontal gyrus BA46</td>
<td>50 36 6</td>
<td>4.80</td>
</tr>
</tbody>
</table>
Fig. 2.1. Brain activations from the comparison of imageable versus non-imageable word pairs (thresholded at p<0.001 uncorrected). (a1) sagittal view highlighting posterior cingulate gyrus, (b1) coronal section highlighting parahippocampal gyri bilaterally; (c1) axial slice highlighting left (red) and right (blue) angular gyri; (d1) axial section highlighting right inferior frontal gyrus; (a2, b2-3, c2-3, d2) plots of parameter estimates at each of activation sites showing how each condition (imageable-semantic=I+S; imageable-nonsemantic=I+NS; nonimageable-semantic=NI+S; nonimageable-nonsemantic=NI+NS) contributes to the main effect. The y-axes show the effect size in arbitrary units.
b) Non-imageability versus Imageability

Main effects when contrasting nonimageable versus imageable conditions, revealed differential activations in left inferior frontal gyrus (BA47), left anterior superior temporal gyr (BA38), left superior temporal sulcus (BA21/22), and right cerebellum (see table 2, fig. 3). From the parameter estimates (fig. 3a2, 3b2-3) it is evident that there are different relative regional contributions to this effect as a function of semantic relatedness. Especially notable is the difference in activation between the two nonimageable conditions in left inferior frontal gyrus. Comparing these two conditions as simple main effects, greater activation to nonimageable nonsemantic, relative to nonimageable semantic, was evident in left inferior frontal gyrus (p<0.001 uncorrected).

Table 2. Brain regions revealed from the comparison of non-imageable versus imageable word paired-associates

<table>
<thead>
<tr>
<th>Brain area</th>
<th>Coordinates</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>L anterior superior temporal gyrus BA38</td>
<td>-58 10 -18</td>
<td>4.73†</td>
</tr>
<tr>
<td>L inferior frontal gyrus BA47</td>
<td>-54 20 6</td>
<td>4.37†</td>
</tr>
<tr>
<td>L superior temporal sulcus BA21/22</td>
<td>-60 -42 4</td>
<td>4.14†</td>
</tr>
<tr>
<td>R cerebellum</td>
<td>26 -82 -32</td>
<td>4.93++†</td>
</tr>
</tbody>
</table>
Fig. 3. Brain activations revealed by comparing non-imageable versus imageable word pairs (viewing threshold is p<0.001, uncorrected); (a1) axial view highlighting anterior superior temporal gyrus, (b1) axial view highlighting inferior frontal gyrus (red) and superior temporal sulcus (blue); (a2, b2-3) plots of parameter estimates for each of the activation sites showing how each condition (imageable-semantic=I+S; imageable-nonsemantic=I+NS; nonimageable-semantic=Nl+S; nonimageable-nonsemantic=NI+NS) is contributing to the observed effect. The y-axes represent the effect size in arbitrary units.

c) Semantic relatedness versus semantic unrelatedness
Another relevant aspect of this study was the contribution of semantic relatedness during encoding of word pairs. Two degrees of semantic relatedness were included in the experimental conditions: close relatedness and unrelatedness. Encoding-related activations for semantically related versus the unrelated paired-associates when compared revealed activations in right precuneus (BA7), right middle frontal gyrus (BA8), right intraparietal sulcus (BA40), posterior cingulate (BA31), right middle temporal gyrus (BA21), and right frontomarginal gyrus (BA10) (see table 3, fig. 4). Several contributions to this main effect is evident in the plots of parameter estimates (Fig. 4b2, 4d2). A notable difference in contribution to the effect from the two semantically related conditions motivated further examination of simple main effects. These revealed a greater contribution to the main effect from imageable
semantically related, relative to nonimageable semantically related, in right middle frontal gyrus and in posterior cingulate (p<0.005, uncorrected).

Table 3. Brain regions revealed from the comparison of semantically related versus semantically unrelated word paired-associates

<table>
<thead>
<tr>
<th>Brain area</th>
<th>Coordinates (x y z)</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>R precuneus BA7</td>
<td>16 -62 36</td>
<td>5.42*†</td>
</tr>
<tr>
<td>R middle frontal gyrus BA8</td>
<td>42 16 42</td>
<td>5.07*†</td>
</tr>
<tr>
<td>R intraparietal sulcus BA40</td>
<td>42 -50 60</td>
<td>5.04*†</td>
</tr>
<tr>
<td>R posterior cingulate gyrus BA31</td>
<td>12 -28 46</td>
<td>4.55†</td>
</tr>
<tr>
<td>R middle temporal gyrus BA21</td>
<td>68 -16 -8</td>
<td>4.39†</td>
</tr>
<tr>
<td>R frontomarginal gyrus BA10</td>
<td>36 54 -4</td>
<td>4.16†</td>
</tr>
</tbody>
</table>
Fig. 4.1. Brain activations revealed by comparing semantically related versus unrelated word paired-associates (viewing threshold is p<0.001, uncorrected). (a1) axial view highlighting precuneus activation; (b1) coronal view highlighting middle frontal gyrus activation; (c1) coronal view highlighting intraparietal sulcus activation; (d1) coronal view highlighting posterior cingulate activation; (a2, b2, c2, d2) plots of parameter estimates for each of the activation sites shows condition specific (imageable-semantic=I+S; imageable-nonsemantic=I+NS; nonimageable-semantic=NI+S; nonimageable-nonsemantic=NI+NS) contributions to the observed effect. The y-axes reveal effect size in arbitrary units.
d) Semantic unrelatedness versus semantic relatedness

The main effect of comparing semantically unrelated versus related word pairs, independently of imageability, revealed activation in left inferior frontal gyrus (BA45), left inferior temporal gyrus (BA20), left posterior cingulate gyrus (BA23/30), right caudate nucleus, and bilateral cerebellum (see table 4, fig. 5). These areas are thus especially involved in making associations between items when there is no existing semantic link between them. Furthermore, examining the parameter estimates (Fig. 5a2-4, 5b2) for the various conditions showed different relative contributions as a function of degree of imageability to the main effect in a subset of regions. Examination of simple main effects revealed a greater contribution from nonsemantic nonimageable, relative to nonsemantic imageable in left inferior frontal cortex (p<0.001, corrected). This implies that this region is especially
involved in making associations between not only items which hold no semantic relation but where that relationship is abstract.

Table 4. Brain regions revealed from the comparison of semantically unrelated versus semantically related word paired-associates

<table>
<thead>
<tr>
<th>Brain area</th>
<th>Coordinates</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>L inferior frontal gyrus BA45</td>
<td>-44 30 -20</td>
<td>5.05*†</td>
</tr>
<tr>
<td>L posterior inferior temporal gyrus BA20</td>
<td>-52 -52 -16</td>
<td>4.41*†</td>
</tr>
<tr>
<td>L posterior cingulate gyrus BA23/30</td>
<td>-8 -48 8</td>
<td>4.17†</td>
</tr>
<tr>
<td>R caudate nucleus</td>
<td>18 10 14</td>
<td>4.12†</td>
</tr>
<tr>
<td>R cerebellum</td>
<td>32 -70 -18</td>
<td>4.72†</td>
</tr>
</tbody>
</table>
Fig. 5. Brain activations revealed by comparing semantically unrelated versus related word pairs (viewing threshold is p<0.001, uncorrected). (a1) axial section highlighting inferior frontal gyrus (red), posterior cingulate (blue) and caudate nucleus (green); (b1) coronal section highlighting inferior temporal gyrus activation; (a2-4, b2) plots of parameter estimates for each of the activation sites showing how each condition (imageable-semantic=I+S; imageable-nonsemantic=I+NS; nonimageable-semantic=NI+S; nonimageable-nonsemantic=NI+NS) contributes to the observed effects. The y-axes represent effect size (arbitrary units).
discussion

The aim of this study was to examine the influence of imagery and semantic relatedness upon the patterns of brain activation during encoding of word paired-associates. Behaviourally, the imageable and semantically related word pairs resulted in better subsequent memory, as evident in highest cued-recall performance. Performance in the imageable semantically unrelated and nonimageable semantically related conditions was also significant relative to chance levels. However, although still above chance, performance for the nonimageable and semantically unrelated condition was significantly lower than for the other conditions, consistent with the proposal that both imagery and semantic relatedness are important mnemonic aids.

Paivio's dual-coding hypothesis (Paivio, 1986), holds that words are represented both verbally and pictorially. Thus, the distinct neural responses arising from comparisons involving the two factors of interest in this study, imagery and semantic relatedness speak to the validity of the dual-coding hypothesis. However, the data also indicate that verbal and nonverbal semantic systems are not completely autonomous but possibly coupled subsystems (Shallice, 1988). For instance, brain activation patterns observed in this study reveal a degree of complementarity between verbal and nonverbal properties in the material encoded. More specifically, these two systems seem to operate in parallel but allowing the most prominent psychological aspects of the encoded information to be the focus of processing.

In terms of the psychological dimension of imageability, which a priori would be expected to engage an imagery system, the most remarkable activation was that of parahippocampal cortex. It is noteworthy that this region has previously been reported as activated during encoding of imageable material (complex colour photographs) in a manner predictive of subsequent memory for the encoded items.
(Brewer et al., 1998). One further study (Krause et al., 1999a) also replicated bilateral parahippocampal activation during encoding of highly imaginable word-pairs which predicted subsequent memory, i.e. successful recall. The present study indicates that this cortical region is involved in high level visual processes necessary for encoding imageable material. Furthermore, a characteristic of activation in this area was further enhancement during encoding of imageable word pairs, which had no semantic association. This suggests that when other processes that support encoding, such as activation of existing associations between imageable items, are not feasible, then a reliance on the use of imagery mediation further augments activation in this region. This finding, in particular, provides additional evidence that processes instantiated in the parahippocampal region during encoding relate to imageability, as has been previously suggested (Wise et al., 2000).

The other region showing a response to imagery was posterior cingulate. It is noteworthy that activation in posterior cingulate was observed in all main comparisons, though in distinct locations, except for nonimageable versus imageable contrast. A possible explanation for the functional significance of different conditions activating different sites in the posterior cingulated is that this area’s connections to many different other brain regions might lead to different parts of this structure being activated when processing different kinds of material. The contrast of imageable versus nonimageable suggests a general contribution to processing imageable material, regardless of semantic content. A related region of posterior cingulate showing an effect of semantic relatedness, shows a further enhancement as a function of imageability. The likely role of this region in imagery is also suggested by previous reports of activation in response to processing concrete (i.e. highly imageable) nouns (Nobre et al., 1997). Thus, posterior cingulate and parahippocampal gyri, which are connected via an association fiber system (Afifi & Bergman, 1998) and receive inputs from visual association cortex (Pandya & Yeterian, 1985), may contribute to the process of attribution of imageability in the service of learning.
Two other regions of activation during encoding of imageable paired-associates were angular gyrus (BA19) and occipitoparietal fissure. The angular gyrus is part of major association cortex in the inferior parietal lobule, which shares connections with all sensory cortical areas, is known to play a role in object recognition (Afifi & Bergman, 1998). These activations might reflect a role for high-level visual association cortices in providing access to visual memory representations required for making visual associations between word pairs. Interestingly, a region of left inferior temporal gyrus, an area with a known role in object recognition (Miyashita, 1993; Chao et al., 1999a) and visual memory storage (Harries & Perrett, 1991), showed activation in response to encoding semantically unrelated but imageable word pairs, when compared to semantically unrelated non-imageable items. It is possible that creating an association between items, which are not inherently associated in a meaningful way, requires a linking representation. What is suggested here is that during encoding of semantically unrelated non-imageable word-pairs, subjects were probably trying to link both items by visualising a linking image and that this process might be mediated by the inferior temporal gyrus.

A common feature of posterior areas, such as those in visual occipital and parietal cortex, which respond to imagery is that they receive inputs from visual cortex (Pandya & Yeterian, 1985), which might explain preferential engagement in processes involving a high-level visual component. Thus, the role of these areas in encoding imageable (concrete) word pairs, or in the generation of a mediatory image between unrelated items, might relate to engaging a pictorial representation that links these words. More generally, the activation of regions with a strong affinity to high-level visual processes provides neuroanatomical support for Paivio's word pictorial representational model within the dual-processing hypothesis (1986).

A prediction from this study is that paired-associates, which are not easily encoded via an imagery strategy, would engage a verbal non-pictorial representational system. This is the second proposition in Paivio's dual coding hypothesis (1986). It is also central to Bower & Anderson's conceptual-propositional hypothesis (1973).
More specifically the prediction is that nonimageable processing during encoding requires neural systems that mediate information manipulation through verbalisation. It is notable that the contrast comparing nonimageable versus imageable word paired-associates revealed activations in regions strongly associated with language functions. Thus, a significant contribution to encoding of nonimageable word paired-associates was found in left superior temporal gyrus and sulcus. These regions have been widely reported as activated in language studies, especially when comparing semantic with phonological processing (Pugh et al., 1997; Price et al., 1997; Binder et al., 1997). Furthermore, language comprehension functions are widely acknowledged to depend on activity in posterior temporal cortex (Wernicke's area) (Wernicke, 1874), close to the region implicated here.

Additionally, several studies have identified lateralised patterns of activation, especially in prefrontal cortex, according to type of material encoded, with left-sided activations being observed during encoding of verbal material and right-sided activations during encoding of visual-based material, e.g. faces (Kelley et al., 1998; Lee et al., 2000). Neuropsychological investigations of nonfluent aphasic patients have also highlighted the importance of left prefrontal cortex in verbal processing (Buckner et al., 1996b). This study replicates a previous finding of activation in left inferior frontal gyrus during cued recall of nonimageable word paired-associates (Fletcher et al., 1996). Furthermore, encoding nameable objects, which can be represented both verbally and pictorially is reported as resulting in bilateral prefrontal activations (Kelley et al., 1998). Such a pattern of lateralisation might also apply when verbal material is either concrete (imageable) or abstract (nonimageable).

In the present study left prefrontal activations correlated with encoding nonimageable word pairs and right prefrontal activations with encoding imageable word pairs. Notably, lesions to left inferior frontal gyrus have been observed to correlate with deficits in generation of semantically appropriate action words for
concrete nouns (Thompson-Schill, 1998). The relevance of this finding is that it suggests a possible role for this brain area in mediating verbal semantic associations by linking nouns with verbs. Additionally, semantic priming experiments have also revealed that activation reductions are circumscribed to inferior prefrontal cortex (Bucker et al., 2000; Wagner et al., 2000).

The other factor manipulated in this study was semantic relatedness. Behaviourally, words containing a high number of possible associates were more memorable, as has previously been shown (Jenkins & Russell, 1952; Deese, 1959), and led to activation of brain systems that are related to semantic organisation (Kapur et al., 1994a, 1994b; Wagner et al., 1998a; Thompson-Schill et al., 1998). The comparison between semantically related versus unrelated word pairs revealed activation mainly in the right hemisphere, involving an extensive region of intraparietal sulcus. This activation was most pronounced in the case of imageable semantically related items. Thus, one possible role in encoding word pairs denoting a high degree of concreteness is likely to be utilisation of representations encoded in other cortical areas, such as those necessary for object recognition and language comprehension areas (Afifi & Bergman, 1998). In fact, it has previously been proposed that in paired associate learning, the concreteness of items might allow a greater number of meaningful associative linkages. In such cases the image of an object might activate semantic associations from the object's parts, and those which are similar either in image or meaning (Watkins & Watkins, 1975). Additionally, these semantic associations might be facilitated in the case of abstract items in the word pairs by activating concrete meaningful linkages related to interactive actions between the two items.

Medial parietal cortex (precuneus) was activated in response to learning semantically related relative to unrelated word pairs, regardless of imageability. Medial parietal cortex has been implicated in retrieval processes, particularly for imageable material (Tulving et al., 1994b; Shallice et al., 1994; Fletcher et al., 1995a; Buckner et al., 1996a; Shallice et al., 1994). A recent study, however,
reported precuneus activation regardless of whether the encoded material is concrete or abstract and visually or auditorily presented (Krause et al., 1999b). One possibility that might explain the present finding is that learning semantically related, relative to unrelated, word pairs might evoke more visual links, regardless of whether the material is actually concrete or abstract. For instance, paper-pen might evoke the image of a writer, or crime-punishment might evoke the image of a prisoner. By contrast, semantically unrelated words, whether concrete or abstract, lack a readily available imageable association, as in lettuce-piano or comfort-fantasy.

Previous studies have suggested putative neural systems engaged in manipulation of information through semantic processing or encoding information according to meaning. For instance, it is well known that in learning experiments randomly organised material tends to be ordered in an "effort after meaning" (Bartlett, 1932). Active retrieval of semantic knowledge is suggested as one of the roles of inferior frontal cortex (Buckner et al., 1995; Demb et al., 1995; Démonet et al., 1992; Kapur et al., 1994b; Martin et al., 1995a; Petersen et al., 1989; Petersen et al., 1990). However, during encoding of semantically related items there is a likely reliance on retrieval of already established relations. In relation to retrieval, a hypothesis has been put forward, which underlies the functional differences between DLPFC and VLPFC. According to the two-level hypothesis, (Petrides, 1994) conscious efforts to retrieve a specific piece of information requires VLPFC, whereas automatic retrieval requires DLPFC.

In the present study a substantial contribution to encoding of word paired-associates was observed in prefrontal cortex. Patterns of activation in these regions might relate to the amount of search required for making connections between words. All comparisons elicited prefrontal activations, which varied in location. For instance, DLPFC (middle frontal gyrus) was activated in response to the semantically related relative to unrelated, conditions, especially when the paired-associates were imageable.
Notably, a distinct pattern of lateralisation emerged within the prefrontal activations, with mainly right-sided prefrontal areas activating during encoding of both imageable and semantically related word pairs and left-sided activations responding to non-imageable and semantically unrelated conditions. Previous studies have reported lateralisation specific to prefrontal cortex as a function of encoding or retrieval (Tulving et al., 1994a; Shallice et al., 1994; Fletcher et al., 1995b; Nyberg et al., 1996b; Grady et al., 1995; Haxby et al., 1996; Oin et al., 1996). For instance, according to the HERA model (Tulving et al., 1994a) left prefrontal cortical activity is mainly related to encoding and right-sided activation to retrieval. Such a lateralisation pattern might reflect the relative contributions from right-sided structures in eliciting existing associations between the study items (retrieval) and from left-sided structures in creating new ones (encoding).

Right-sided DLPFC activations were found in relation to encoding imageable and semantically related word pairs. This pattern might correlate with monitoring retrieval of already stored representations and associations between items (cued retrieval). In fact, several studies have suggested that right DLPFC could be implicated in monitoring information retrieved from episodic memory (Shallice et al., 1994; Henson et al., 1999b), and in mediating monitoring processes necessary for optimal recall (Fletcher et al., 1998b).

In contrast, VLPFC activation was expressed in the case of semantically unrelated, especially nonimageable, word pairs, suggesting active search for paired-associates, which inherently contained no semantic linkage. A suggested role for VLPFC in active retrieval is depth-of-processing theory (Craik & Lockhart, 1972), which supports that inferior prefrontal cortex contributes to processing items "deeply", by means of semantic processing (Kapur et al., 1994a). Additionally, it has been shown that left inferior prefrontal cortex activations are generally present in verbal tasks but more specifically when the task involves semantic processing (Kapur et al., 1994b; Demb et al., 1995). In this case, there may be semantic processing facilitating new semantic linkages between items normally unrelated.
Left prefrontal cortex has previously been implicated in predicting subsequent memory of verbally encoded material during a semantic decision task (Wagner et al., 1998c). More specifically left inferior gyrus has been shown to be involved in semantic encoding and on-line retrieval of semantic information (Demb et al., 1995), in tests of semantic association (Ricci et al., 1999), and in the process of verb generation in the context of concrete nouns (Thompson-Schill et al. 1998). A possible mechanism to create interactive associations between two nouns is verb generation by assignment of action between the two items. Such a process involves manipulation of verbal aspects of the paired-associates by means of accessing semantic information to match the two items.

**Conclusion**

In summary, the more imageable and semantically related items are, the more memorably they will be associated. The neuroimaging results from this study underlie the functional neuroanatomical differences in brain systems subserving various processes required to make material of different degree of imageability and semantic relatedness memorable. In line with Paivio's dual encoding theory, this study provides evidence for the participation of specific brain structures in encoding words when there is reliance mainly on their verbal representations when pairing abstract items, as well as when pictorial representations are readily evoked by pairing concrete items. In general, a greater engagement of posterior visual association areas was observed during concrete imagery-based processing, and more temporal language areas during abstract verbal-based encoding. Additionally, encoding paired-associates with different degrees of semantic relatedness led to different patterns of activation. Prefrontal areas were primarily activated during processes requiring creation or modification of semantic connections. A dorso-ventral dissociation was noted, where DLPFC was recruited during the process of retrieving connections of pre-existing semantic associations (automatic retrieval) and VLPFC during search for new semantic links (active retrieval). Also within prefrontal regions, lateralisation patterns were observed.
Namely, right PFC seemed more involved in manipulating already existing associations, and left PFC in establishing associative links between unrelated items.

Overall, all types of word pairs, whether concrete or abstract, were supported both by imagery and verbal strategies, with different emphases according to degree of concreteness or abstractness. This finding is very much in tune with Paivio's dual coding hypothesis (1986) that words are represented verbally as well as pictorially and with the hypothesis that verbal and nonverbal (e.g. visual) semantic systems might be considered coupled subsystems (Shallice, 1988). In this study, the neural systems supporting these two types of information processing indicate separation as well as overlap during learning word paired associates. Thus, these results contribute to the description of neural networks involved in verbal associative learning under different mnemonic rules. It also provides neuroanatomical support for Paivio's dual theory. In future studies it would be useful to focus on delineating the possible changes that might take place in these networks over time as a consequence of learning.
The neural changes associated with learning have been studied principally at the molecular and cellular levels (e.g. long-term potentiation, Bliss & Lømo, 1973; Bliss & Collingridge, 1993) where an underlying principle of memory formation is change in neural connection strength (Hebb, 1949). Behaviourally, learning can be defined as the acquisition of new information, which is manifest as increased fluency on a range of psychological measures. Learning necessarily involves change over time and consequently time is critical to its definition. In this fMRI experiment the goal was to characterise time-dependent neural changes associated with associative learning across sensory modalities at the systems level.

Neuroimaging studies of memory have used multiple psychological paradigms to study different aspects of learning (see review by Cabeza & Nyberg, 1997). However, few studies have examined time-dependent aspects of learning. Notable exceptions are studies investigating practice effects on patterns of brain activation which distinguish between brain areas involved in unskilled effortful performance and those related to skilled effortless performance (Friston et al., 1992; Jenkins et al., 1994; Raichle, 1994; Jueptner et al., 1996; Petersen, 1998). For instance, Raichle et al. (1994) used a verbal response selection task to compare a naive with a practised state. Brain regions, including anterior cingulate, left prefrontal, left posterior temporal cortices, and right cerebellum, became less active with practice, with a reverse pattern being observed in bilateral sylvian-insular cortex and left medial extrastriate cortex (Raichle et al., 1994). A more recent study, also using a verbal response selection task, has provided additional evidence of a practice-
related shift in neuronal response (Petersen et al., 1998). In this case early activations in left prefrontal, anterior cingulate and right cerebellar hemisphere were distinct from later activations observed after practice, which were principally located in sylvian-insular cortex.

A major assumption in memory research is that psychologically distinct memory systems have different neuroanatomical components (Baddeley, 1996; McCarthy & Warrington, 1990). The components of these systems differ according to type of material processed and the operations necessary for effective learning. Hence, verbal learning is believed to involve primarily bilateral prefrontal cortex, posterior cingulate, precuneus, and superior temporal gyrus (Grasby et al., 1993; 1994) while visual learning involves extrastriate and higher order visual areas in the occipital and temporal cortices (Tovee et al., 1996; Eacott & Heywood, 1995). It needs to be noted that all these studies have investigated learning associations between stimuli and responses in conceptually related domains. Petrides (1985) has argued that brain systems involved in processing stimuli and responses, which have no intrinsic relationship to each other, are different to those brain systems recruited during learning associated stimuli.

The major interest in the present study was to investigate the neural systems that come into play during the process of learning between arbitrary nonverbal visual and auditory stimuli. The presupposition in this study was that learning is best characterised in terms of temporal changes and consequently the focus of the present study was on time-dependent effects on neural response observed during associative learning. The experiment involved a discrimination task where auditory and visual stimuli were presented in pairs under two different conditions, consistent and inconsistent pairings. The prediction was that learning relationships between arbitrary pairs of stimuli, regardless of sensory modality, would activate left medial temporal and right frontal regions. This prediction is derived from previous work showing that left medial temporal lobe activity is associated with learning and memory processes (Petrides, 1985b; Wallenstein et al., 1998; Gabrieli, 1998;
Eichenbaum, 1992; Fletcher et al., 1995b; Henke et al., 1997; Kelley et al., 1998; Rombouts et al., 1997; Squire & Zola, 1996). Additionally, right prefrontal activations have been implicated in incremental learning during repeated trials (Kopelman et al., 1998). Differences in the networks activated in response to consistent or inconsistent audiovisual pairs were also expected, since learning consistent associations across repeated presentations involves retrieval of knowledge of pair combinations, whereas inconsistent associations require further attentional demands due to the continuous novelty of combinations of presented items. Consistent pairs are thus more likely to activate regions involved in retrieving encoded information, such as previously reported right frontal and precuneal areas (Tulving et al., 1994b; Shallice et al., 1994; Fletcher et al., 1995a; Buckner et al., 1996; Kopelman, 1998). However, with attentional demands being greater for inconsistent audiovisual pairings, involvement of visual and cingulate areas is likely to increase, consistent with findings that anterior cingulate activates with enhanced task demands (Pardo et al., 1990; Posner & Petersen, 1990; Kosslyn et al., 1996; Benedict et al., 1998; Carter et al., 1998). Additionally, activations in superior right frontal and medial temporal cortices were predicted in conditions where two unrelated stimuli had to be associated, based on previous findings that implicate these regions in associational tasks (Petrides, 1985b; Wallenstein et al., 1998).

**methods**

**Subjects**

14 healthy volunteers (5 females) whose ages ranged from 22 to 29 years participated in the study. All were free of neurological or psychiatric illness. Nine subjects took part in a behavioural study while the remaining five (all right-handed) took part in the functional neuroimaging experiment. All subjects provided informed consent. This study was approved by the joint Ethics Committee of the National Hospital for Neurology and Neurosurgery and the Institute of Neurology, London.
Experimental materials

The auditory stimuli consisted of 6 distinct musical chords. The visual stimuli were 6 distinct and novel Chinese ideograms. By “novel” is implied that subjects were not familiar with the stimuli and thus hold no previous semantic representations of them. Half of the auditory and the visual stimuli were presented as consistent audiovisual pairs, i.e. always the same two items in the pair. The other half were presented as inconsistent audiovisual pairs, i.e. the same two items in the pair were never presented paired repeatedly within a single training session. The auditory and visual stimuli were presented in pairs, either consistently or inconsistently, in two thirds of trials while in the remaining third of trials the same 12 auditory and visual stimuli were presented unpaired.

The six musical chords were constructed with a Sound effects program by adding together the single components according to the frequencies assigned to single notes (http://www.afx.com/Wlips/hm000026.htm). The Chinese ideograms were obtained from (http://www.wg.omron.co.jp/cgi-bin/j e/nocolor/dict).

Cognitive task

Subjects were studied in an event-related paradigm during three separate training sessions. During these periods, subjects were exposed to four different conditions: consistent audiovisual pairs, inconsistent audiovisual pairs, single auditory stimuli, and single visual stimuli. Behavioural data were acquired outside the scanner due to technical problems which prevented the acquisition of behavioural data during scanning. Hence, nine different subjects underwent exactly the same experimental procedure as that involved in the neuroimaging sessions. Three test sessions interleaved with the study sessions provided the behavioural data, which was subsequently analysed using a Wilcoxon paired test.

Subjects were informed that they would receive auditory and visual stimuli either paired or unpaired. They were instructed to learn whether audiovisual pairs were
consistently or inconsistently paired, and to ignore the unpaired stimuli. No behavioural response was required during the study sessions. During test sessions, also present in the neuroimaging task, subjects had to make button responses in the same fashion that behavioural subjects were instructed to do outside the scanner. The task requirement was to press a key when they thought a presented pair belonged to the consistent condition and another key if it belonged to the inconsistent condition.

**Experimental Procedure**

Subjects lay in a MRI scanner for approximately 30 minutes. Auditory stimuli were presented through headphones at a comfortable volume regulated individually to each subject. Visual stimuli were presented in white, against a black background, and centred on a screen located approximately 20 cm away from the eyes, subtending a visual angle of 7°.

The total presentation duration for each stimulus pair was 3 s. For presentation of audiovisual pairs, the auditory stimulus was presented for 3 s and the visual stimulus overlapped for the last second (see Fig. 1).

![FIG.1. Graphic representation of the temporal order and duration of presentation of stimuli.](image)

The three stimulus types (paired consistent, paired inconsistent, and unpaired) were randomly presented in triplets and each event type was repeated three times. The inter-stimulus interval, which was randomly jittered (±1/3 and ±2/3 TR) was 12.9 s on average. There were 9 repetitions of each paired event and 12 repetitions of the single events, per session, with critical differences in presentation
for the 9 events belonging to each paired condition at each session. Specifically, three consistent audiovisual pairs were repeated three times, amounting to a total of 9 presentations, whereas 9 different inconsistent pairs were presented once but these were rearrangements of three different possible item pairings. This arrangement was necessary to keep the number of event presentations equal while creating a pattern of consistency by repetition in the consistent condition, and a pattern of inconsistency, by avoiding repetition in the inconsistent condition. In total, over the entire experiment, there were 27 repetitions for events corresponding to the consistent condition, 27 repetitions for events corresponding to the inconsistent condition, 18 repetitions of single auditory stimuli, and 18 repetitions of visual stimuli.

**Magnetic Resonance Procedures**

The data were acquired from a 2-Tesla Magnetom-VISION whole-body MRI system (Siemens, Erlangen) equipped with a head volume coil. Multislice T2* weighted fMRI images were obtained with a gradient echo planar sequence and axial slice orientation (echo time TE=40 s, repetition time TR=4.3 s, 64x64x48 voxels). Data were acquired in three sessions. A total of 279 contiguous volume images were acquired. Each echoplanar image comprised 48 1.8 mm slices taken every 3 mm, positioned to cover the whole of the cerebrum. Each session comprised 93 volume images.

Structural images were obtained in the same orientation, yielding T1 weighted images with 1x1x1.5 mm voxel size (matrix size: 256x256x108). An MPRAGE (magnetised prepared rapidly acquired gradient echo) sequence was used, with 108 partitions acquired in the sagittal plane, with a flip angle of 12°, and TE=4 ms, TR=9.7 ms, TI=600 ms. Structural images were acquired before the start of the experiment, followed by one or two scout scans and five dummy scans, before the first functional scan, to allow for T2 equilibration effects. The functional scans were acquired in an event-related fashion. The scanner was synchronised with the
jittered onset of events, and the ratio of interscan to interstimulus interval ensured that voxels were sampled at different phases relative to stimulus onset (a total of three scans were taken per event). This allows the gathering of data to take place at different stages of the haemodynamic response and also minimises expectation effects.

**Data analysis**

The data were analysed using the general linear model, in this instance linear regression on condition-specific waveforms convolved with a haemodynamic response function (Friston *et al.*, 1995a) as implemented in SPM98 and SPM97devel. More specifically, the time series were realigned and resliced using a sinc interpolation, adjusting for residual motion-related changes, spatially normalised (Friston *et al.*, 1995b; Friston *et al*. 1996) to a standard EPI template (Evans et al., 1992, 1993; Collins et al., 1994), and smoothed spatially with a 10 mm FWHM isotropic Gaussian kernel and temporally with a 4 s Gaussian kernel, and a high-pass filter providing a basis set of cosine functions with a cut-off period of 61 s to remove low frequency drifts in the BOLD signal (Holmes *et al*., 1997).

A group analysis was performed using a fixed-effects model. The effects included in the design matrix were 15 sessions each corresponding to three sessions per subject (3 sessions, 5 subjects) including separate columns for main effects, time by condition interactions, mean values for each condition, and low-pass filter values. The effects of interest involved four conditions: consistent audiovisual pairs, inconsistent audiovisual pairs, auditory unpaired stimuli, and visual unpaired stimuli. Paired contrasts tested for greater learning effects in one versus another condition both averaged over time (main effects) and as a function of time (time-dependent changes in activation). The time by condition interactions were computed within session. Thus, the results show activations which increased over time during all individual sessions.
Auditory unpaired stimuli were used to compute differential main and time-dependent effects between paired and unpaired material. The reason why this control was used and not the visual unpaired control is that the duration of presentation of the auditory unpaired stimuli was equal to the presentation duration of both paired conditions, whereas the visual unpaired control was much shorter, matching only for the presentation duration of the visual stimuli in the paired conditions (Fig. 1). Time-dependent effects were modelled as linear increases in BOLD signal, such that a comparison between A and B would predict increasing response over time for condition A and the reverse for condition B.

* A priori activations were predicted on the basis of previous studies of encoding and retrieval in episodic memory. The regions reported here were those which reached correction for voxel-wise multiple comparisons (p<0.05*) or those selected on the basis of pre-specified hypotheses where arbitrary thresholds of cluster size ≥40 voxels and height p<0.001 uncorrected were involved in order to avoid type II errors.

---

**results**

**Behavioural data**

Nine subjects were tested outside the scanner following three exposures to the material emulating the scanning conditions (see Fig. 2). Thus, the procedure was similar to that during scanning, with alternate tests after each learning session for the purpose of measuring learning rates. The test consisted of presentations of audiovisual pairs where the subjects had to press the right button on a keypad if they identified the pair as consistent, or the left button if they identified it as inconsistent.
However, the device which reports button presses from within the scanner room was not functioning during the scanning sessions for this experiment. Thus, subjects were debriefed after the experiment and according to their reports, they all learnt the associations appropriately. In order to compensate for the shortage of behavioural data, the same experiment was carried out on 9 other subjects outside the scanner, from which the actual level of learning was inferred.

### Behavioural performance

![Graphic plot of the behavioural results obtained from the test performed. The first, second, and third columns (white, grey, and black) indicate the mean group performance for each session when tested for (a) both paired conditions (consistent and inconsistent; white); (b) consistent audiovisual pairs (grey); and (c) inconsistent audiovisual pairs (black). The difference in mean group performance between the first and the third sessions is significant for all conditions (Wilcoxon paired test p<0.01).](image)

The fact that 12 items were presented at test, three of which were consistent and nine of which were inconsistent, will incur in a difference in chance levels for each condition, namely chance in the consistent condition is 25%, whereas in the inconsistent condition it is 75%. However, two-sample tests showed a significant increase above chance in correct responses between the first and the third testing sessions (Wilcoxon paired p<0.01) both for the consistent and the inconsistent conditions separately.
Functional Neuroimaging data

The effects of interest in this study are changes in neural response for consistent (where behavioural data indicate significant learning) compared to inconsistent audiovisual pairs across time i.e. in association with learning, as well as changes in neural response for the paired conditions versus the auditory unpaired condition.

a) Main effects

Paired versus Unpaired conditions
Time-independent main effects were found in response to learning audiovisual pairs as compared to processing unpaired unimodal stimuli. When comparing both paired conditions (consistent and inconsistent) versus the single auditory control condition, foci of activation were localised bilaterally in the occipital lobe (with peak activation reached in right inferior occipital gyrus \[x,y,z = 34,-76,-18\], \(Z\)-score=7.32, \(p<0.05\) corrected, Fig. 3). The observed effect reflects greater response for the paired conditions, which was differentially larger to that observed for the auditory control condition.

b) Time-dependent interactions

This analysis contrasted differential neural responses as a function of time, where the time scale reflected each individual session. In other words, within-session time-by-condition interactions were computed in order to characterise differential changes in response over the duration of individual sessions.
Time-dependent BOLD-type changes for consistent versus inconsistent audiovisual pairs

This contrast revealed significant time-dependent differential responses in left medial parietal cortex (precuneus) and right dorsolateral prefrontal cortex (see Table 1, Fig. 4). These are regions where there was a significant differential response between consistent and inconsistent pairs. In both cases (precuneus, Fig. 5-left and dorsolateral prefrontal cortex, Fig. 5-right) the increased signal was
greater for the consistent condition relative to the inconsistent condition, although
the difference became less significant over time.

The signal changes illustrated in figures 5, 7, and 8 should be interpreted as
follows: the changes correspond to each condition within each session and they
are relative to one another, i.e. where the error bars stand far apart, the difference
in increase of activation was bigger between conditions than when the error bars
almost overlap each other. So, for instance, the first graph in figure 5 (left
precuneus) implies that the increases in activation for the consistent condition (red
line) were greater than for the inconsistent condition (blue line) within the first
session and this difference decreased over time. The fact that left precuneus
emerged as a significant time-by-condition interaction is then probably driven by
the significant difference observed during the first session.

Table 1. Local maxima for areas whose activation increased over time
for the consistent versus the inconsistent audiovisual pairs (p<0.001
uncorrected).

<table>
<thead>
<tr>
<th>Brain location</th>
<th>Coordinates (x y z)</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left precuneus (BA7)</td>
<td>-10 -64 60</td>
<td>4.30</td>
</tr>
<tr>
<td>Right dorsolateral prefrontal (BA46)</td>
<td>48 28 24</td>
<td>3.34</td>
</tr>
</tbody>
</table>

FIG.4. (a) Coronal section highlighting dorsolateral prefrontal cortex and (b) axial section highlighting
precuneus activations when comparing time by condition interactions between the consistent and the
inconsistent audiovisual pairs, thresholded at p<0.001 for viewing.
FIG. 5. Graphs illustrating signal change at each session as mean parameter estimates in precuneus (left) and dorsolateral prefrontal cortex (right) when comparing time by condition interactions between the consistent and the inconsistent audiovisual pairs (a.u. = arbitrary units). The responses to inconsistent do not significantly go up over time. The effect shown here is that, as the data were modelled as increases in activation over time, the difference between the consistent and the inconsistent conditions was greatest at the beginning, when the associations were new, and decreased over time as the subjects presumably had learnt the associations.

**Time-dependent changes for inconsistent versus consistent audiovisual pairs**

The opposite contrast revealed significant time-dependent differential responses for inconsistent relative to consistent audiovisual pairs in right angular gyrus, left middle and inferior occipital gyri, right cuneus, anterior cingulate, and right cerebellum (see Table 2, Fig. 6). In all extrastriate (Fig. 7) and cingulate (Fig. 8) areas these differential responses were greater for the inconsistent pairs than for the consistent pairs. What is meant by this contrast is that where there is a difference between the two conditions, for that session increases in activation were greater for inconsistent than for consistent. Note that the time-by-condition interactions reported here were within session, i.e. the increases take place at each session, and are relative to each other (e.g. inconsistent relative to consistent).
These differences, illustrated in Figs. 7-8, show distinct patterns of modulation as a function of time. Each data point reflects the amount of signal change at each session, and the plots reflect responses to one condition relative to another, e.g. whether the difference in response between the two conditions of interest increases or decreases between conditions over time. In the cases of left middle occipital (BA18) and right cuneus (BA18) response differences for both paired conditions are reduced over time. In left middle occipital gyrus (BA19/37) and left inferior occipital gyrus (BA18) the differences become more marked over sessions. In the case of anterior cingulate, the time-dependent differential response in BA32 is due to greater response for the inconsistent relative to the consistent paired condition mainly during the second session (Fig 7-left), whereas in BA24 there is a gradual decrease in differential response with only the first session showing a significant difference between conditions (Fig. 8-right).

Table 2. Local maxima for areas whose activation increased over time for the inconsistent versus the consistent audiovisual pairs (p<0.001 uncorrected).

<table>
<thead>
<tr>
<th>Brain location</th>
<th>Coordinates (x y z)</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right angular gyrus</td>
<td>56 -50 22</td>
<td>4.24</td>
</tr>
<tr>
<td>Left middle occipital gyrus (BA18)</td>
<td>-28 -84 12</td>
<td>3.67</td>
</tr>
<tr>
<td>Left middle occipital gyrus (BA19/37)</td>
<td>-46 -70 -8</td>
<td>3.27</td>
</tr>
<tr>
<td>Left inferior occipital gyrus (BA18)</td>
<td>-38 -88 -4</td>
<td>3.14</td>
</tr>
<tr>
<td>Right cuneus (BA18)</td>
<td>22 -104 4</td>
<td>3.05</td>
</tr>
<tr>
<td>Left anterior cingulate (BA32)</td>
<td>-10 48 0</td>
<td>3.17</td>
</tr>
<tr>
<td>Right anterior cingulate (BA24)</td>
<td>14 4 40</td>
<td>3.09</td>
</tr>
<tr>
<td>Right cerebellum</td>
<td>4 -74 -22</td>
<td>3.50</td>
</tr>
</tbody>
</table>
FIG. 6. (a, c) Lateral and (b) back rendered views of the brain regions activated (red=right angular gyrus, green=extrastriate cortex, blue=anterior cingulate) when comparing time by condition interactions between inconsistent and consistent audiovisual pairs, thresholded at p<0.001 for viewing.

FIG. 7. Plots of signal change (mean parameter estimates) for each session in extrastriate cortex resulting from the time-dependent comparison between inconsistent and consistent paired conditions (a.u. = arbitrary units).
FIG. 8. Plots of signal change (mean parameter estimates) for each session in cingulate cortex resulting from the time-dependent comparison between inconsistent and consistent paired conditions.

Time-dependent changes for paired versus unpaired conditions

Time-dependent effects of processing associations between stimuli in different sensory modalities were compared to processing within a single modality, namely the unpaired auditory stimuli. Comparing both paired conditions against the auditory unpaired condition (Table 3) revealed a differential effect in left posterior hippocampus (Fig. 9). The effect observed in left posterior hippocampus reflects a differential enhanced response to the paired conditions relative to the auditory controls, where this difference is attenuated over time (Fig. 10).

Table 3. Areas of activation modulated by time for consistent and inconsistent audiovisual pairs versus the single auditory stimuli (p<0.001 uncorrected; *p=0.001)

<table>
<thead>
<tr>
<th>Brain location</th>
<th>Coordinates (x y z)</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consistent Pairs – Auditory Unpaired</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left hippocampus</td>
<td>-28 -44 2</td>
<td>3.20*</td>
</tr>
<tr>
<td><strong>Inconsistent Pairs – Auditory Unpaired</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left hippocampus</td>
<td>-28 -42 0</td>
<td>4.26</td>
</tr>
</tbody>
</table>
c) Modulatory response in extrastriate cortex

A further contrast compared how signal changed over time for both the visual and auditory control conditions. The motivation for this contrast was to investigate whether single auditory stimuli could evoke visual responses when they are predictive of a visual stimulus with learning (i.e. over time). For this purpose the two single control conditions were compared, namely the single auditory and the single visual stimuli. These stimuli were also used in the paired conditions, thus it can be expected that with learning of the audiovisual associations, presentations of single auditory stimuli might evoke a predictive effect of its paired visual stimulus. This predictive effect was also found to have a neuroanatomical basis, namely single presentations of auditory stimuli were observed in extrastriate cortex, localised to right fusiform gyrus (maxima=38, -74, -20). This area also activated to single presentations of visual stimuli but there was a marked difference in the modulation of this response over time for single auditory and visual stimuli. The response in this area remained fairly constant for the visual control with a slight decrease over time possibly due to repetition effects, whereas there was a noted increase in response for single auditory stimuli (Fig. 11). This modulatory effect is investigated in depth in chapter 6.
In this study it was hypothesised that learning to associate nonverbal auditory and visual stimuli would involve operations such as sensory processing as well as establishing and monitoring whether auditory-visual stimuli are consistently or inconsistently paired. Repeated presentations of consistent, but not inconsistent, pairs probably were expected to produce automatic recognition of the auditory stimuli, presented first, and to become predictive of the subsequent presentation of a specific visual stimulus.

**Comparisons between paired audiovisual conditions**

Learning to discriminate consistent from inconsistent pairs requires formation of specific associations between auditory and visual stimuli. As subjects learn to recognise the consistent pairs, presentation of preceding sounds provides a context for retrieval of the associated visual stimuli. A contrast comparing the main effects of both paired conditions versus the auditory unpaired control revealed activations bilaterally in the occipital lobe extending to precuneus. These activations might reflect response differences to the kind of material being processed in both sides of the comparison. In particular, the visual activations are
likely to reflect the fact that audio-visual pairs are compared to auditory stimuli alone.

Since learning is by definition a dynamic process occurring over time, the main focus in interpreting the results from this study is in brain activations reflecting time-dependent changes. The learning process involved in this experiment is likely to comprise a dynamic interface between encoding and retrieval operations, with encoding being more dominant in the initial stages and retrieval more prevalent in later stages. In this regard, it is notable that brain regions activated over time included medial parietal and right dorsolateral prefrontal cortex. These regions have been widely reported as part of a retrieval network for either auditory or visual stimuli alone (Tulving et al., 1994b; Shallice et al., 1994; Fletcher et al., 1995a; Buckner et al., 1996). It might be the case that in learning consistent audiovisual pairings, retrieval strategies gradually start operating concurrently, or subsequently, to encoding operations. This suggestion is supported by the finding that re-presentation of previously presented material leads to right prefrontal cortex and precuneus activations (Kopelman et al., 1998).

In terms of functional specification, right prefrontal cortex is implicated in retrieval processes involving success, effort, or attempt (Wagner et al., 1998; Schacter & Buckner, 1998; Buckner et al., 1998a, 1998b; Rugg et al., 1996). In this experiment, the success hypothesis is a possible explanation since time-related changes in performance correlate with the amount of material retrieved. More specifically, Rugg et al.'s (1998) suggestion, namely that right anterior prefrontal cortex is involved in evaluating whether retrieved information in response to a test item represents an appropriate prior episode, is consistent with the present finding. In addition, the right dorsolateral prefrontal cortex has also been linked to monitoring of information retrieved from episodic memory (Shallice et al., 1994; Henson et al., 1999b), and to mediating monitoring processes necessary for optimal recall (Fletcher et al., 1998b). This putative monitoring role for the dorsolateral prefrontal cortex applies particularly to the current paradigm where
subjects assess response success over time and each repetition of consistent pairs allows confirmation of prior correct or incorrect responses.

Additionally, Kelley et al. (1998) have reported laterisation within the dorsal prefrontal and the medial temporal cortices according to the nature of the material encoded, such that right prefrontal activations were found during encoding of nonverbal material. Results pertaining to this study might support this suggestion in so far as the material used was nonverbal and prefrontal activation was right-sided. However, a critical difference is that in the current study right prefrontal activation reflects learning-related changes rather than steady state conditions.

The other time-dependent differential pattern of activation was seen in left medial parietal cortex. The medial parietal cortex is associated with visual imagery, and more specifically with cued recall, rather than encoding, of imageable verbal paired associates (Grasby et al., 1993, 1994; Fletcher et al., 1995b; Mellet et al., 1996; Platel et al., 1997). In this case medial parietal cortex activation increased over time paralleling increasing familiarity for the consistent pairs. With time there is an increase in the degree to which presentation of auditory stimuli automatically cues the associated visual stimulus which, due to its very nature (Chinese ideogram), cannot be verbalised by non-Chinese speakers. Thus, this activation might reflect processes necessary for generating a visual image of ideograms in the consistent condition. Note that in the inconsistent condition the cue, i.e. the sound, cannot generate a specific visual image as there is no consistent visual stimulus pairing.

During learning, subjects increasingly distinguish inconsistent from consistent pairs, a task that requires generation of hypotheses in relation to whether the pairs were consistent. Such hypotheses could be subsequently revised (Elliot and Dolan, 1998) with practice and learning. It is likely that attentional demands in the inconsistent condition were greater than in the consistent condition based on the fact that a greater number of audiovisual combinations were presented. In this study, time-dependent relative activations in two anterior cingulate foci, ventral and
dorsal, were observed for the inconsistent condition relative to the consistent condition. Anterior cingulate is a component of a network responsible for selective attention regardless of sensory modality (Posner and Petersen, 1990; Benedict et al., 1998), as is the angular gyrus (Kosslyn et al., 1996), the other principle focus of activation. It is likely that these activations in conjunction with those observed bilaterally in extrastriate cortex are due to greater attentional demands under the inconsistently paired condition.

The more dorsal cingulate region observed has in previous studies been implicated in selection between competing alternatives (Pardo et al., 1990; Elliot and Dolan, 1998) and in error detection (Carter et al., 1998). In the present task, subjects must make a prediction of whether the pairs are consistent or inconsistent, and this process is more demanding for inconsistent trials. An animal study shows that lesions in a region corresponding to the presently described ventral focus of the cingulate lead to failure of learned response extinction (Roberts et al., 1992). A similar process might be taking place in this study, where initial hypotheses regarding inconsistent pairs being consistent are subsequently rejected.

Analysis of time-dependent changes controls for non-specific effects by the use of a reference state (consistent versus inconsistent and vice versa). The legitimacy of these comparisons rests on an assumption that non-specific effects, such as time-dependent effects, are equally expressed in both states. This possibility of differential interaction between task and non-specific factors cannot be totally outruled and should caution interpretation of these findings. Furthermore, other non-specific interaction effects may be found underlying the difference of frequency in item presentation for consistent versus inconsistent conditions during each session. These possible interactions are not likely to override the main results presented and it might be noted that differential interactions between task and non-specific factors is limited (Petersson et al., 1999).
Time-dependent changes for paired versus unpaired conditions

These contrasts highlight differential time-dependent within-session responses for bimodal associative, as opposed to single-item, processing. The main differential response for both, paired, compared to unpaired, conditions, was in left posterior hippocampus. These time-dependent changes reflect greater increase in response for the paired conditions relative to the auditory unpaired condition, especially during the first session. The fact that these differential effects decline over sessions might imply that recruitment of this area is maximal when the material is novel (Knight, 1996; Dolan & Fletcher, 1997; Xiang & Brown, 1998; Gooding et al., 2000). The fact that both kinds of pairs show similar time-dependent responses may reflect commonalities specific to associative processing across two sensory modalities.

An extensive animal and human literature highlights the critical role of the hippocampus in associative learning (Bunsey & Eichenbaum, 1996; Eichenbaum, 1992; Henke et al., 1997; Rombouts et al., 1997). A more recent proposal is that the hippocampus is related to associative processes for discontiguous items in space or time (Levy et al., 1983, 1996; Wallenstein et al., 1998). The hippocampal responses observed in this study are consistent with both these suggestions. Note that for both consistent and inconsistent conditions, the task involves not only items being presented in different sensory modalities but also in a temporal dimension where auditory and visual stimuli appeared at different points but always separated by the same interval. From this perspective, the differential response for bimodal relative to unimodal conditions fits with neuropsychological findings that left hippocampal lesions lead to impairments in nonspatial conditional association tasks (Petrides, 1985b), where subjects have to make associations between two items presented one shortly after the other. The current task shares many similarities with Petrides's task (1985). In Petrides's study, frontal-lobe and left temporal-lobe patients (including the left hippocampal region) were impaired in a nonspatial associational task involving learning associations between six coloured-
light caps and six hand postures matched in unique one-to-one combinations. These patients were able to discriminate between single stimuli but had difficulty in selecting the correct movement to a given visual stimulus.

The hippocampal response observed in the present study was located in posterior hippocampus. A functional rostro-caudal gradient along the hippocampus has been suggested in a meta-analysis of Lepage et al. (1998), who proposed a model (HIPER model) claiming that encoding in the hippocampus is reflected primarily rostrally and retrieval caudally. Recent empirical data also provides support for this suggestion (Strange et al, 1999). However, Schacter & Wagner (1999), in a review of fMRI studies, suggest that posterior medial temporal lobe is associated with episodic encoding, whereas the anterior part might play a role in relational encoding. The posterior hippocampal differential response reflecting changes with time would seem to support the retrieval-based proposal in so far as retrieval processes are increasingly engaged with learning. Moreover, the left medial temporal region has previously been found to activate during incremental learning, i.e. repetitive trials, a process which might involve consolidation of new memories (Kopelman et al., 1998).

**Conclusion**

The present results shed light on the neuranatomical substrates underlying associative learning, and confirmed the previously demonstrated role of hippocampus in this function. However, further controls are required to derive any conclusions regarding crossmodal associative learning. Since the comparisons in this study focused on crossmodally paired items versus single items, further comparisons involving crossmodal versus intramodal pairs are required. This approach will constitute the basis of the following chapter. Additionally, this study contains the caveat that the neuroimaging data are not related directly to behavioural measures of learning gathered from the subjects who underwent
scanning but that these were provided by another set of subjects. This latter point should be carefully considered in future studies.
In the previous chapter I described time-dependent changes in neural response during learning where associations were made between auditory and visual stimuli. Notably, areas identified in bimodal learning are part of a well-documented retrieval memory system for material presented within a single modality. Additionally, time-dependent responses in posterior hippocampus during associative processing were suggestive of a role for this region in forming links across sensory modalities.

However, the nature of the controls used in the previous study, namely single auditory and visual stimuli, while providing a reasonable comparison for the associative processes taking place in the paired consistent and inconsistent audiovisual conditions, do not provide a sufficient basis to make clear inferences about crossmodal associative learning. In order to show associative processes specific to audiovisual pairs, i.e. associations between different sensory modalities, further controls are required, where associations are established between items within the same sensory modality. This approach constitutes the basis of the study presented in this chapter, where audiovisual crossmodal pairs were compared to visual and auditory intramodal pairs.

A role for human hippocampus in learning has been widely reported (Grady et al., 1995; Tulving et al., 1994c; Stern et al., 1996; Fernández et al., 1999; Teng & Squire, 1999; Eichenbaum 1999; Barrash et al., 2000; Rolls, 2000b). However, most learning studies employ stimuli in a single modality. Several exceptions to this generalisation are exemplified by studies investigating odour-place associative
learning in rats (Lipton et al., 1999) and neuropsychological observations in patients with medial temporal lobe lesions of impairment in nonspatial conditional associational tasks involving learning arbitrary associations between stimuli and responses (Petrides, 1985b). A possibility has been raised that hippocampus plays a role in recognising temporal contingencies in associative learning (Wallenstein et al., 1998), which constitutes a crucial aspect of the present study.

"Crossmodal learning", the process of linking two items to form an associated representation, might be mediated by convergence of stimuli presented in different sensory modalities onto brain sites that support multisensory integration. Potential candidate structures mediating this convergence have previously been proposed (Ettlinger & Wilson, 1990; Hadjikhani and Roland, 1998). For instance, the insula-claustrum complex is known for its wide connectivity with the rest of the brain (Sherk, 1986; Augustine, 1996) and consequently it is proposed as a mediator of communication between different modality-specific perceptual/memory systems (Ettlinger & Wilson, 1990). In humans, this region has been implicated in crossmodal transfer between different sensory modalities, for example touch to vision (Hadjikhani and Roland, 1998). Another candidate region is the superior colliculus, where electrophysiological studies report convergence of different sensory streams onto multisensory cells (Stein & Meredith, 1993; King & Palmer, 1985; Frens & van Opstal, 1997; Wallace et al., 1996), with a suggested role in attending and orienting to different sensory inputs (Sprague & Meikle, 1965; Schneider, 1969; Casagrande et al., 1972; Goodale & Murison, 1975; Stein 1984; Sparks, 1986; Stein & Meredith, 1990; Meredith et al., 1992).

The main prediction in the present study is a replication of posterior hippocampal activation during associative learning. The modification of the design from the previous study was intended to provide a more precise functional description of the brain regions involved in crossmodal learning. In particular, the use of conditions in which learned associations can be formed between stimuli within the same
modality provides a basis for disambiguating activations attributable to crossmodal sensory integration from those attributable to non-specific associative learning.

**methods**

**Subjects**

13 healthy volunteer subjects (4 females), whose ages ranged from 22 to 30 years, participated in the study. All were free of neurological or psychiatric illness. All subjects provided informed consent. The study was approved by the joint Ethics Committee of the National Hospital for Neurology and Neurosurgery and the Institute of Neurology, London.

**Experimental materials**

The auditory stimuli consisted of 27 distinct sounds, 9 were used in the audiovisual session and 18 in the auditory-auditory session. The visual stimuli were 27 distinct Chinese ideograms; 9 were used in the audiovisual session and 18 in the visual-visual session. Half of the auditory and the visual stimuli were presented as consistent pairs, i.e. pairs always containing the same two items. The other half were presented as inconsistent audiovisual pairs, such that the same two items in the pair were never presented more than once in the same combination. This was the case for all three types of pairs: crossmodal (av), visual intramodal (vv) and auditory intramodal (aa). The sounds were constructed using a multimedia software package (Imsi Masterclips 300,3000) and transformed with a Sound effects program. The Chinese ideograms were obtained from http://www.wg.omron.co.jp/cgi-bin/j e/nocolor/dict.
Cognitive task

Subjects were studied in an event-related paradigm during three separate training sessions, an audiovisual session, an auditory-auditory session, and a visual-visual session. During the audiovisual session, subjects were exposed to three different conditions: consistent audiovisual pairs, inconsistent audiovisual pairs, and null events. During the audio-audio session, subjects were exposed to three different conditions: consistent auditory-auditory pairs, inconsistent auditory-auditory pairs, and null events. During the visual-visual session, subjects were exposed to three different conditions: consistent visual-visual pairs, inconsistent visual-visual pairs, and null events (see Table 1). The order of sessions was counterbalanced across subjects.

Subjects were informed that they would receive auditory and visual stimuli either crossmodally (audiovisual) or intramodally (visual-visual, auditory-auditory) paired. For each training session they were instructed to learn which pairs were consistent (paired in the same combination) or inconsistent (paired in a different combination at each presentation). The task requirement for all sessions (crossmodal and intramodal) was to press a key when subjects thought a pair belonged to the consistent condition and another key if it belonged to the inconsistent condition. Behavioural data, as a measure of learning, were acquired during scanning by recording subjects’ responses on the keypad.

Table 1. Experimental conditions

<table>
<thead>
<tr>
<th></th>
<th>Crossmodal (av)</th>
<th>Intramodal (vv)</th>
<th>Intramodal (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistent</td>
<td>avcon</td>
<td>vvcon</td>
<td>aacon</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>avincon</td>
<td>vvvincon</td>
<td>aalincon</td>
</tr>
</tbody>
</table>
Experimental Procedure

During neuroimaging the auditory stimuli were presented through headphones at a comfortable volume regulated individually to each subject and the visual stimuli were presented in white, against a black background, and centred on a screen located approximately 20 cm away from the eyes, subtending a visual angle of 7°.

The total presentation duration for each stimulus pair was approximately 3 s. The first stimulus in the pair was presented for approximately 1.5 s and the second stimulus, which followed immediately after, without temporal overlap, lasted 1.5 s (see Fig. 1).

<table>
<thead>
<tr>
<th>Pair presentation time: 3 s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
</tr>
<tr>
<td>~1.5 s</td>
</tr>
</tbody>
</table>

FIG. 1 Graphic representation of the temporal order of item presentation for each paired condition (audio-visual, auditory-auditory, and visual-visual).

The three stimulus types, for each session, were randomly presented in triplets (blocks of three), with each event being repeated six times over the whole session. The inter-stimulus interval, which was randomly jittered (+375 ms, +750 ms, 1125 ms, 1500 ms), was 6 s on average. Jittering the presentation of stimuli prevented expectation effects. Also, fixing the onset of stimuli presentation to different points during scans allows data acquisition during different phases of the haemodynamic response.

In total, over each session, there were 18 repetitions for events corresponding to a consistent condition (i.e. 6 repetitions of three pairs), 18 repetitions for events corresponding to the inconsistent condition (i.e. 18 presentations of 6 sounds and 6 images arranged differently every time), and 18 repetitions of null events (blank screen and silence). The duration of the whole experiment was approximately 20 min, involving approximately 6 min per session.
Magnetic Resonance Procedures
The data were acquired from a 2-Tesla Magnetom-VISION whole-body MRI system (Siemens, Erlangen) equipped with a head volume coil. Multislice T2* weighted fMRI images were obtained with a gradient echo planar sequence and axial slice orientation (echo time TE=40 s, repetition time TR=3 s, 64x64x48 voxels). Data were acquired in three sessions. A total of 357 contiguous volume images were acquired from each subject over the whole experiment. Each echoplanar image comprised 32 2mm slices separated by 0.75 mm, positioned to cover the whole of the cerebrum. Each session comprised on average a similar number of volume images.

The functional scans were acquired in an event-related fashion. The scanner was synchronised with the jittered onset of events, and the ratio of interscan to interstimulus interval ensured that voxels were sampled at different phases relative to stimulus onset. This latter procedure allows the gathering of data to take place at different stages of the haemodynamic response and also minimises expectation effects. One or two scout scans and five dummy scans were acquired before the first functional scan, to allow for T2 equilibration effects. Structural images were acquired, at the end of the experiment, in the same orientation, yielding T1 weighted images with 1x1x1.5 mm voxel size (matrix size: 256x256x108). An MPRAGE (magnetised prepared rapidly acquired gradient echo) sequence was used, with 108 partitions acquired in the sagittal plane, with a flip angle of 12°, and TE=4ms, TR=9.7ms, TI=600ms.

Data analysis
The data were analysed using the general linear model, in this instance linear regression on condition-specific waveforms convolved with a haemodynamic response function (Friston et al., 1995a) as implemented in SPM99. More specifically, the time series were realigned and resliced using a sinc interpolation method, adjusting for residual motion-related changes, spatially normalised
(Friston et al., 1995b; Friston et al. 1996) to a standard EPI template (Evans et al., 1992, 1993; Collins et al., 1994), and smoothed spatially with a 8mm FWHM isotropic Gaussian kernel, a high-pass filter providing a basis set of cosine functions with an automatically specified cut-off period of 512 s to remove low frequency drifts in the BOLD signal (Holmes et al., 1997) and a Gaussian filter of 4 s to remove high frequency drifts.

A first-level analysis consisted of creating contrast images for each condition and for each subject. The covariates of interest included in the design matrix corresponded to main effects and time by condition interactions pertaining to each condition of interest, i.e. consistent and inconsistent pairs in each of the three sessions (crossmodal (av), and intramodal (vv and aa)) for each of the 13 subjects. A subsequent second-level analysis involved one-sample t-tests where contrast images produced at the first-level stage were combined to determine group effects using a random effects model. Time-by-condition interactions were modelled as exponentially decaying responses over the duration of the whole experiment, i.e. between sessions. The reason for modelling the data as an exponential decaying function was because the learning taking place in this study was expected to take place mostly during the first part of the experiment, followed by a slight decrease and/or plateuing of the rate of learning over time.

A further manipulation of the data was performed in order to test whether activations circumscribed to a specific region coincided with activation sites from chapter 4. This step consisted in creating a mask image from the contrast which yielded activation in left posterior hippocampus in the previous study, and using this mask in order to carry out a small volume correction in the contrast image which reveals activation in a critical area of interest in the present study.

Reported activations surviving correction for voxel-based multiple comparisons were thresholded at p<0.05. P-values of a priori predicted activations on the basis of previous studies of associative learning and crossmodal integration were
reported individually for each activation site of interest (see tables) and thresholded at $p<0.001$, uncorrected.

results

Behavioural data

Thirteen subjects provided a measure of learning by indicating (through a button-press) whether the pair presented belonged either to the consistent or the inconsistent condition. Subsequently the mean percentage of correct responses was calculated (Fig. 2), revealing significant learning (performance above chance, $>50\%$) across subjects for all three conditions: crossmodal (audiovisual pairs) and intramodal associations (visual-visual and auditory-auditory pairs). Paired t-tests showed a significant ($p<0.05$) difference in mean group test scores, in all cases involving improved performance between the first and last sessions for all conditions.

![Correct responses (%)](image)

**FIG.2.** Graphic plot of learning rates. The x-axis represents the responses in the different conditions: crossmodal audio-visual (X), and intramodal auditory-auditory and visual-visual (I). The data were split into results pertaining to learning consistent (con) and inconsistent (incon) pairs. The y-axis represents percentage of correct responses. To simplify data plotting the group scores were averaged for every two repetitions. The z-axis represents the different repetitions ($r$) of the material over the duration of the experiment; with $(1+2)$ representing mean values obtained for the first and second repetitions of the material; $(3+4)$ representing mean values obtained for the third and fourth repetitions; and, $(5+6)$ representing the fifth and sixth repetitions. Each bar corresponds to mean group values.
Functional Neuroimaging data

The effects of interest in this study are changes in neural response due to learning crossmodal, compared to intramodal, associations. The data analyses examined for main effects (mean activations during each session) and time-by-condition interactions (exponential increases or decreases in response over the course of the study). In both cases the contrasts of interest compared neural responses due to processing crossmodal versus intramodal pairs, and consistently paired stimuli versus inconsistently paired stimuli.

Although main effects were found, they did not reach corrected significance (p<0.05). The focus of the results and discussion is therefore on the time-dependent changes in neural response. These latter effects, which allow observation of time-dependent modulatory responses, are arguably a better reflection of dynamic processes, such as learning (Büchel & Dolan, 2000), where the time scale of learning for each session was ~6min. What constitutes “learning” here is the neural processes underlying the acquisition of new information, which can be behaviourally tested.

a) Crossmodal and Intramodal associations

Crossmodal versus Intramodal

When contrasting changes in activation over time between the crossmodal (audio-visual) and intramodal (auditory-auditory and visual-visual) conditions regardless of consistency of pairing, significant differential responses were found in right posterior insula-claustrum (Table 2, Fig. 3a) and right anterior hippocampus (Table 2, Fig. 3b).
Table 2. Local maxima for areas whose activation was modulated over time for the crossmodal versus the intramodal pairs.

<table>
<thead>
<tr>
<th>Brain location</th>
<th>Coordinates (x y z)</th>
<th>Z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>R posterior insula-claustrum</td>
<td>32 -6 -6</td>
<td>4.37</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>R anterior hippocampus</td>
<td>18 -12 -12</td>
<td>3.72</td>
<td>p&lt;0.001</td>
</tr>
</tbody>
</table>

FIG.3. Responses observed as a result of comparing exponential time effects due to learning crossmodal pairs versus intramodal pairs (regions of interest marked by red arrows): axial sections of the right posterior insula-claustrum (z = -6) (a) and coronal section highlighting right anterior hippocampus (y = -12) (b). The SPM was thresholded at p<0.001 uncorrected for display.

Examination of the parameter estimates for each of the conditions revealed an effect in right posterior claustrum driven mainly by a response decay over repetitions for crossmodal pairs, which was more prominent in the case of crossmodal inconsistent pairs. This contrasted with an increase in response over time for intramodal pairs, which was more prominent for the intramodal consistent pairs (see Fig. 4). In the case of the right anterior hippocampus the pattern of change in activation over time consisted of a general decreasing response to crossmodal pairs, especially to crossmodal inconsistent pairs, which contrasted with a modest increase in response to intramodal pairs (see Fig. 5).
FIG. 4. Graphic plot of percentage of signal change (as parameter estimates—arbitrary units) for time effects of interest in right posterior insula-claustrum when comparing crossmodal (consistent and inconsistent) to intramodal (consistent and inconsistent) conditions. Values on the x-axis are indicative of time with 1 representing mean values obtained for the first and second repetitions of the material; 2 representing mean values obtained for the third and fourth repetitions; and, 3 representing mean values obtained for the fifth and sixth repetitions. Note the main difference in response is between the crossmodal and intramodal conditions at the beginning of the experiment, such that responses from the crossmodal conditions decrease over time and responses from the intramodal conditions increase over time.

FIG. 5. Graphic plot of percentage of signal change (as parameter estimates—arbitrary units) for time effects of interest in right anterior hippocampus resulting from comparing crossmodal (consistent and inconsistent) and intramodal (consistent and inconsistent) conditions. Values on the x-axis are indicative of time with 1 representing mean values obtained for the first and second repetitions of the material; 2 representing mean values obtained for the third and fourth repetitions; and, 3 representing the fifth and sixth repetitions. Note the most remarkable effect here is a decrease in activation for the crossmodal inconsistent condition.
Intramodal versus Crossmodal

When contrasting intramodal versus crossmodal paired conditions, significant time-dependent differential responses were observed in left posterior hippocampus (Table 3, Fig. 6a) and right superior colliculus (Table 3, Fig. 6b). Based on previous findings, these areas were a priori hypothesised to play a role in learning (Rombouts et al., 1997; Wallenstein et al., 1998), or in processing crossmodal information (Stein & Meredith, 1993; Hadjikhani & Roland, 1998; Gonzalo et al., 2000).

In the case of the left hippocampal activation, examination of parameter estimates revealed a time-dependent change, driven by an increasing response across time to crossmodal pairs, especially crossmodal inconsistent pairs, with little change in response to intramodal pairs (Fig. 7.1). Since a similar pattern of hippocampal activation was also found in the previous study (chapter 5), it was necessary to determine whether the locations of activation were the same. For this purpose, two mask images were created at p<0.05 uncorrected from the contrasts yielding left posterior hippocampal activation in the previous study, i.e. consistent audiovisual pairs versus auditory controls, with activation peaking at x,y,z= -28,-44,2 and inconsistent audiovisual pairs versus auditory controls at x,y,z= -28,-42,0. Subsequently a small volume correction (SVC) on the contrast image yielding left posterior hippocampal activation in the present study was performed using these masks to constrain the search volume. For the peak activation in left posterior hippocampus for the present study (x,y,z=-26,-40,8), the SVC, using any of the two masks, at the voxel-level was significant at p<0.05 uncorrected (Z-score=2.04) and at the cluster-level (spatial extent) p<0.001, corrected.

For the superior colliculus, the profile of time-dependent change is driven by an increase in response to crossmodal pairs, especially inconsistent ones (Fig. 7.2). Responses to crossmodal (consistent and inconsistent) and intramodal (consistent and inconsistent) conditions were similar during the first session, with differences in response between these pairs of conditions becoming greater over time.
Table 3. Local maxima for areas whose activation was modulated over time for the intramodal versus the crossmodal pairs.

<table>
<thead>
<tr>
<th>Brain location</th>
<th>Coordinates (x y z)</th>
<th>Z-score</th>
<th>p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>L posterior hippocampus</td>
<td>-26 -40 8</td>
<td>2.04</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>R superior colliculus</td>
<td>6 -26 -6</td>
<td>2.07</td>
<td>&lt;0.005</td>
</tr>
</tbody>
</table>

FIG. 6. (a) Sagittal view of left posterior hippocampus (red); (b) and partial coronal section of right superior colliculus.

FIG. 7.1. Graphic plot of percentage of signal change (as parameter estimates – arbitrary units) for time-dependent effects in left posterior hippocampus, resulting from comparing intramodal and crossmodal conditions. Values on the x-axes are indicative of time with 1 representing mean values obtained for the first and second repetitions of the material; 2 representing mean values obtained for the third and fourth repetitions; and, 3 representing the fifth and sixth repetitions. Note that the most remarkable effect observed here is an increase in response for the crossmodal inconsistent condition.
b) Consistent and Inconsistent associations

The data were also examined to determine differential activation patterns in terms of consistency and inconsistency of pairings, that is pairs of stimuli presented always in the same combination or never in the same combination. When analysing the interaction between crossmodality/intramodality and consistency/inconsistency of pairing (i.e. crossmodal consistent versus crossmodal inconsistent contrasted against intramodal consistent versus intramodal inconsistent), a significant effect was seen in right inferior frontal gyrus (BA45) (xyz= 60 14 4; Z=5.05*, p<0.05 corrected, Fig. 8b). This interaction specifically addresses the question of where in the brain there are greater response increases or decreases over time to crossmodal consistent (and not inconsistent) relative to intramodal consistent (and not inconsistent) pairs. The comparison was modelled
as an exponential function for a relative response decay over time for the crossmodal consistent and intramodal inconsistent conditions and as response increase over time for the crossmodal inconsistent and the intramodal consistent conditions. Examination of parameter estimates revealed that this effect in right ventrolateral prefrontal cortex was driven mainly by an increasing differential response over presentations for crossmodal inconsistent paired stimuli relative to all other conditions (Fig. 8a).

![Graph](image)

**FIG.8.** (a) Plots of percentage of signal change (as parameter estimates -arbitrary units) for the time effects for each condition in right inferior frontal gyrus (BA45). Values on the x-axis are indicative of time with 1 representing mean values obtained for the first and second repetitions of the material; 2 representing mean values obtained for the third and fourth repetitions; and, 3 representing the fifth and sixth repetitions. Note the most notable effect here is an increased in activation for the crossmodal inconsistent condition over time; (b) rendered view of a template brain highlighting the location of activation. Note again that the model is fitting for a relative change across conditions.

The reverse comparison (intramodal consistent versus intramodal inconsistent compared to crossmodal consistent versus crossmodal inconsistent) revealed a response in left precuneus (Table 4, Fig. 9). Other foci of differential activation, similar to those reported in the previous chapter, were localised to the cingulate region bilaterally and to right angular gyrus (Table 4, Fig. 10). This interaction addresses the question of where in the brain responses increase or decrease more over time specifically to intramodal consistent (and not inconsistent) than to
crossmodal consistent (and not inconsistent) pairs. The data for this comparison were also modelled as an exponential response decay over time for the intramodal consistent and crossmodal inconsistent conditions and as response increase over time for the intramodal inconsistent and the crossmodal consistent condition. Fig. 10 presents the pattern of activation changes described in precuneus over time, with the effect being driven by increasing activation in response to consistent associations and decreasing activation in response to inconsistent associations. The difference in response between crossmodal consistent pairs and crossmodal inconsistent pairs decreases over time.

Table 4. Local maxima for areas whose activation was modulated over time in response to an interaction between intramodal vs. crossmodal and consistent vs inconsistent conditions.

<table>
<thead>
<tr>
<th>Brain location</th>
<th>Coordinates (x y z)</th>
<th>Z-score</th>
<th>p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>L precuneus</td>
<td>-4 -66 34</td>
<td>3.71</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>R posterior cingulate sulcus</td>
<td>10 -44 42</td>
<td>4.45</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>L posterior cingulate gyrus</td>
<td>-4 -26 40</td>
<td>4.15</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>R angular gyrus</td>
<td>38 -50 36</td>
<td>4.20</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

FIG.9. Rendered view of a template brain highlighting the activation loci in precuneus (blue arrow), posterior cingulate (purple arrows) and angular gyrus (magenta arrow).
FIG. 10. Graphic plot of percentage of signal change (as the parameter estimates—arbitrary units) for the time effects of interest for each condition in left precuneus. Values on the x-axis are indicative of time with 1 representing mean values obtained for the first and second repetitions of the material; 2 representing mean values obtained for the third and fourth repetitions; and, 3 representing the fifth and sixth repetitions.

discussion

The aim of this experiment was to follow up the findings from the previous study (Chapter 4), in order to identify neural systems specifically involved in crossmodal associative learning. The preceding study lacked the necessary controls to justify any firm conclusions regarding key activations, reflecting crossmodal learning. The present paradigm allowed the comparison between learning crossmodal (in the auditory and visual modalities) and intramodal (with both stimuli in the pair being either auditory or visual) stimuli pairs. The limitation encountered in the previous study was that the controls used allowed comparisons only between crossmodal pairs and single stimuli, which informs strictly about the process of forming associations between two items relative to processing non-associated items. Thus, it was not possible to conclude that the previous findings objectively distinguished between generic associative learning from more specific formation of crossmodal
associations. However, one of the findings from the prior study, namely left posterior hippocampal involvement in associative learning, is replicated in the present study, suggesting a common neural system underlying both generic associative learning and specific crossmodal associative learning.

The design used in this study was similar to the previous study in that pairs of stimuli were divided into consistent and inconsistent pairings. However, on this occasion pairs belonging to each of the three conditions (audio-visual, audio-audio, and visuo-visuo) were presented repeatedly in different sessions, to avoid overloading subjects with different types of material in one single session. The behavioural data show a significant difference in performance between the first and last presentations of the material for each of the three conditions, indicating that neural responses correlated with processes of associative learning. Since the emphasis of the study is on learning, the discussion focuses primarily on the results pertaining to time-dependent modulations of response for the crossmodal and intramodal associations.

A note of caution is also here important regarding interpretation of increases and decreases in activation since they often refer to the conditions in question and not to some absolute change in activation for a specific condition. That means that an increase in activation resulting from a contrast might represent an increase in activation of condition A relative to a decrease in B, an increase in B and no change for B, or no change for A relative to a decrease for B. In future studies it might be sensible to ask questions about the significance of the nature of the changes in activation based on well specified hypotheses regarding the direction of change in activation. A valid a priori hypothesis of decreases in activation is, for instance, that a task becomes practiced, familiar, less novel, less difficult, thus requiring fewer neural resources, in the form of e.g. reduced oxygen consumption. The equivalent would apply in the case of increases in activation, i.e. requirement of greater neural resources when the task is new or difficult, such as in the initial stages of learning.
**Insula-claustrum**

In a comparison of the differential time-dependent responses engendered by processing crossmodal relative to intramodal pairs, activation was evident in insular cortex, localised to the right posterior insula-claustrum region. This suggests that this structure is involved in processes specific to crossmodal association formation, and adds new information to the results replicated from the previous study (chapter 4). Several studies to date have suggested a possible role for the insula-claustrum region in crossmodal processing (Hadjikhani and Roland, 1998; Webster et al., 1993; Ettlinger & Wilson, 1990; Sherk, 1986). The connectivity of the claustrum is well characterised, receiving and giving rise to cortical projections, and containing maps of different sensory (visual, auditory, and somatosensory) and motor systems (see review by Sherk, 1986). Generally, the strategic location of the insula-claustrum complex explains its wide connectivity with the rest of the cortex, which has been described in detail elsewhere (see reviews by Augustine, 1996 and Sherk, 1986). This extensive network of connections converging on this structure might explain its putative role in processes involving integration between more than one sensory modality. In fact, such a mechanism for crossmodal processing has been proposed, whereby ventral claustrum mediates communication between different modality-specific perceptual/memory systems (Ettlinger & Wilson, 1990).

Human neuroimaging studies have also provided evidence for a role in crossmodal processing in this region. A PET study of crossmodal transfer in humans reported claustrum activations in response to a condition where transfer occurred between the tactile and visual modalities (Hadjikhani and Roland, 1998). Right insula activations have also been reported in the context of a PET study of synaesthesia, in this case passive listening to words which evoked colour sensation in synaesthetes (Paulesu et al., 1995). Thus, the insula-claustrum region is a possible neuroanatomical candidate for mediating multisensory convergence. This implies that allowing convergence of inputs in the insula-claustrum might be an important
part of the integration process. In fact, the insula-claustrum activation in this study resulted from AV>AA+VV.

The evidence suggests a role for the insula-claustrum in crossmodal processing. It could be that this structure is specifically involved in mediating inputs from different sensory modalities and/or from unimodal cortices, and therefore it becomes activated in processes where various modalities are integrated, such as in the present crossmodal associative paradigm. In the current study, the time-dependent effect observed in this region was mainly driven by a decrease in response over time to crossmodal, relative to intramodal, pairs. It is possible that posterior insula-claustrum is involved in allowing convergence of inputs from different modality-specific sensory cortices, and that this gradual reduction of activation over time reflects greater efficiency of this operation, as crossmodal associations are consolidated over repeated presentations.

**Hippocampus**

A relevant aspect of all associative processes is timing of stimulus presentation. The temporal relationship between two stimuli will determine the strength of the contingency, regardless of the modality of presentation. Contiguous presentation of two stimuli, relative to separate presentations in time, facilitates recognition of specific matchings. With repeated pairings, presentation of a first stimulus (cue) in a pair triggers an expectation of the ensuing stimulus. This proposal derives from simple instance-based models which emphasise an identification advantage for familiarised instances, which can be observed after just one previous presentation of an item (Schacter, 1987a). In the case of the present study, similar mechanisms might be taking place, with the difference being that the familiarised instance is composed of two related items. Thus, hippocampus might contribute to the process of learning to identify compound presentations of contiguous stimuli as familiar unified paired representations, and inconsistently paired items as unfamiliar pairings.
Neuroimaging studies have reported a learning role for human hippocampus (Grady et al., 1995; Tulving et al., 1994c; Stern et al., 1996; Dolan & Fletcher, 1997). Hippocampal involvement in learning has been identified mainly through use of material involving one study modality. In the present study, two hippocampal regions, one in left posterior hippocampus and another in right anterior hippocampus, are implicated in the process of associative learning involving items presented in different sensory modalities. The left posterior hippocampus response (peaking at $x,y,z=-26,-40,8$) replicates that found in the previous study ($x,y,z=-28,-44,2$), such that masking the latter with the former contrast image, revealed a positive overlap in the location of activation. In both studies subjects were instructed to learn about audiovisual associations, either in consistent or inconsistent pairings, where there was learning for consistent pairings as a function of increasing familiarity with repeated presentations, whereas inconsistent pairings retain a degree of novelty given that its components are always presented in different combinations.

The differential effect in left posterior hippocampus becomes more evident over time. This effect consisted of a relative increasing activation over time in response to presentation of crossmodal inconsistent pairs. The relative differential increase in activation over time in response to crossmodal inconsistent pairings might be due to continuing novelty of these pairs given that they are continuously rearranged with every presentation. The notion that left hippocampus may be involved in novelty detection has previously been suggested, for instance in the case of change detection in category-exemplar pairings (Dolan & Fletcher, 1997). However, the second effect observed in the case of the crossmodal consistent pairs, i.e. initial increase followed by decrease in activation, might reflect an initial learning effect, which is reduced with repetition and increasing familiarity to the material (Strange et al., 1999).

In right anterior hippocampus, the pattern of response observed over time consisted of decreasing activation to inconsistent (especially crossmodal) pairings,
and, to a lesser extent, increasing activation to consistent (especially intramodal) pairs. A possible interpretation for this pattern of response might be that as the material becomes more familiar with repetition, a greater predictability develops in response to the initial item in the pair which serves as a cue for the ensuing item. It is thus conceivable that for the consistent conditions predictability increased with time as the pairs were repeated, whereas for the inconsistent conditions predictitability decreased with time as the inconsistent pattern of cue and paired stimuli presentation became apparent.

These findings may not necessarily be representative of a specific hippocampal role in the formation of associations between stimuli in different sensory modalities. Rather, they add to existing evidence that this brain structure is involved in associative processes, including stimulus-stimulus relations, which have not been widely explored, such as in the case of crosmodal associations. It is also worth noting that several studies have demonstrated hippocampal activity during learning across different domains. For instance, odour-place association learning has been shown in rodents (Lipton et al., 1999), and, in humans, lesions encroaching medial temporal lobe are known to impair learning nonspatial conditional associations between arbitrary stimuli and responses (Petrides, 1985b). In general terms, these present findings strengthen the idea that hippocampus plays a general role in associative learning, independent of material type.

This argument regarding the hippocampal involvement in associative learning can be further advanced. In many learning experiments, stimuli have a semantic value, which allows previous knowledge or associations to facilitate the encoding of familiar material in a new arrangement. In such cases, brain areas involved in semantic processing are likely to participate in learning. However, the brain is also capable of encoding new information or associations, which are completely devoid of pre-specified meaning. This capacity to create associations between disparate, novel, items might also be supported by the hippocampus. The degree of novelty in this case would diminish with time, as subjects become familiar with the new
material. However, if the repetitions of the material consist of new associations of the same material, a certain degree of novelty will be maintained. More specifically, the pattern of lateralisation observed for these hippocampal activations suggests involvement of lateralised differences for processing novelty (left hippocampus) and for matching encoded material with subsequent presentations of the same material (right hippocampus).

These two different patterns of activation in hippocampus observed in the present study seem to be driven by a response to the crossmodal inconsistent pairs, relative to the other conditions (see figs. 5 and 7.1). This condition exemplifies the case of having the same stimuli presented in ever-changing arrangements over the course of the session. The pattern of activation observed in left posterior hippocampus is one of a gradual decreased response with time to crossmodal inconsistent pairs, whereas the reverse pattern is observed in right anterior hippocampus. A more elaborate explanation for this differential response might be that the left hippocampus is processing the novelty component of the varying associations, which does not fade over time, whereas right hippocampus might be processing the degree of novelty of the single stimuli in the associations, which does fade over time. This reduction in novelty might correlate with greater recognition of these single stimuli in the pairs, maybe by continuous comparison of current stimuli with those stored from previous presentations. This interpretation implies that left hippocampus is more involved in the associative aspects of presenting two stimuli as a pair, and right hippocampus in responding specifically to matching present stimuli with past presentations.

**Superior colliculus**

Creating associations between stimuli in different sensory modalities is likely to involve brain structures which are normally engaged in processing unimodal inputs (sensory cortices) and, in addition, other structures receiving inputs from unimodal processing sites, which facilitate the process of crossmodal association. Insula-
claustrum might be one of these sites, as has been previously discussed. However, electrophysiological studies in the mammalian brain have suggested another mechanism of integration whereby different sensory streams converge onto individual ‘multisensory’ cells. Such type of integration has been observed in superior colliculus (Stein & Meredith, 1993; King & Palmer, 1985; Frens & van Opstal, 1997; Wallace et al., 1996). Since this structure receives visual, auditory, and somatosensory inputs (Stein, 1984), from unimodal sensory cortices, convergence of these different types of information might be facilitated. Another relevant aspect of the integrative functions of superior colliculus is processing of temporal information. Multisensory interactions can be maximised by overlapping the periods of peak activity of the unimodal discharge trains. This might be achieved by simultaneous presentations or, in some instances, by non-overlapping but sequential presentation.

A tentative explanation of superior colliculus involvement in the present task relates to attending to selected stimuli presented in pairs and in different sensory modalities and orienting behavioural responses as the specific associations were being encoded. The pattern of time-dependent change in response in this area is not very clear, with an overall tendency to increasing activation over time in response especially to inconsistent crossmodal pairs. The lack of significance of this interaction renders any interpretation on the actual role of this region in this specific process speculative.

**Other activations**

Another aspect of interest in this study is the possible differentiation between formation of consistent pairings versus learning to recognise stimuli presented in different combinations, and thus leading to inconsistent associations. I consider the former process to be akin to associative learning, as is conceptualised by virtue of two stimuli, which appear presented in some temporal, spatial or conceptual contingency pattern. The latter process involves greater effort since the natural
tendency is to retain associations as processed during the first presentation. Thus, learning to recognise the stimuli as inconsistent while being presented in different pairings requires a process of hypothesis reversal and updating at each new presentation. For this more demanding task, greater attentional demands were reflected in various extrastriate (precuneus and angular gyrus) and posterior cingulate activations. These findings partially replicate those from the previous chapter, where the relevance of these areas is further discussed.

Additionally, ventrolateral prefrontal cortex activation was observed as a result of processing inconsistent pairs of stimuli with increasing activation over time for crossmodal pairs and decreasing activation in the case of intramodal pairs. Ventrolateral prefrontal cortex has previously been associated with associative functions (Passingham et al., 2000) and is known to receive inputs from different sensory cortices, thus facilitating associations between different sensory modalities. A similar area, just above left inferior frontal sulcus, was reported to activate in response to a task where the encoded material had a very low degree of organisation, i.e. where maximum effort was required from subjects to establish organisation to facilitate encoding (Fletcher et al., 1998a). This finding is coherent with the present observation of increasing activation over time in ventrolateral prefrontal cortex in response to crossmodal inconsistent audiovisual pairs, where the task requirements were much harder than when the pairings were consistently arranged. The increase in activation might reflect an increased ability to predict which auditory cues were paired inconsistently to different visual stimuli, a more effortful process than learning to predict the consistent pairings. The difference in side of activation might be due to different types of material given that in Fletcher et al.’s (1998a) study the stimuli were verbal, whereas in this experiment sounds and pictures were used. Another study suggested a role for right ventral prefrontal cortex in externally cued retrieval (Fletcher et al., 1998b), which is a strong component in the present study, by virtue of individual presentations of a cueing stimulus being followed either by a consistent associate or an inconsistent, random match. Moreover, Petrides (1994) has also put forward a theory, which makes a
functional distinction between ventral and dorsal prefrontal cortex, whereby the former engages preferentially in active search or conscious retrieval effort. This might certainly be the case in the task pertaining to this study, since the presentation of the cue may induce subjects to actively recall its possible associate, which in the case of the inconsistent condition might involve greater retrieval efforts due to the lack of reinforced pairing.

In some cases, one could argue that the activations resulting from contrasting crossmodal versus intramodal pairs are merely related to difficulty of the task – i.e. it is harder to do the crossmodal than unimodal, and thus these areas might be more modulated over time as learning eventually occurs. However, in the case of the consistent pairings at least, the behavioural data yielded no evidence that the crossmodal pairings were more difficult than the intramodal ones. And a t-test comparing the crossmodal inconsistent versus intramodal inconsistent behavioural data points over time also revealed no significant (p>0.05) difference.

**Conclusion**

The evidence from these data supports the view of current models of hippocampal function in associating stimuli which hold some kind of contingency, in this case one of temporal contiguity. Additionally, a lateralisaton pattern in hippocampal activation was noted with a possible relative involvement in processing novelty (left hippocampus) and in matching items with familiar/existing representations (right hippocampus). The main question addressed in this study was whether hippocampus was specifically involved in crossmodal associative learning, as opposed to merely general associative processes regardless of presentation modality. The present results are suggestive of a role for hippocampus in crossmodal associative learning, particularly in the case of inconsistently paired items, where novelty is a crucial factor. The other major finding was the involvement of the insula-claustrum complex in learning associations across modalities, possibly by mediating convergence of inputs from unimodal cortices,
allowing items, which would otherwise be processed individually, to be clustered as unified or linked representations. Notably, the insula-claustrum activation revealed in this study was due to a response enhancement, i.e. the response observed during learning of crossmodal pairs was greater than that of the summed responses of both intramodal learning conditions. Future studies should consider the possible consequences of presenting multiple items consecutively or simultaneously on the neural operations required for associative learning.
CHAPTER 6: Cortical Modulation in Crossmodal Associative Learning

introduction

A traditional perspective is that visual and auditory cortices are tuned to selectively respond to visual and auditory input, respectively (Rakic, 1988; Levitt et al., 1997). However, the brain may also exhibit a high degree of plasticity, which can result in experience-dependent reorganisation in an otherwise normally connected brain. Such reorganisation can be induced artificially by, for instance, effecting crossmodal rewiring at some critical point during development, allowing experience to strengthen the new connections. Animal studies have provided evidence that cortical plasticity can be achieved by a variety of manipulations (Sur et al., 1990; Roe et al., 1990, 1992; Gao & Pallas, 1999; Sharma et al., 2000; vonMelchner et al., 2000). For example, neonatal diversion of retinal axons to the auditory thalamus results in primary auditory cortex resembling visual cortex in its response properties and topography (Sur et al., 1990; Roe et al., 1990, 1992; Gao & Pallas, 1999; Sharma et al., 2000; vonMelchner et al., 2000). Furthermore, it has been observed that cells in primary visual cortex of a normally developed cat brain can be activated by nonvisual stimuli (Bental et al., 1967; Spinelli et al., 1968). Similarly, plasticity in the auditory cortex of guinea pigs can be induced by classical conditioning, whereby receptive field responses to specific frequencies can be trained to increase or decrease, following a brief learning experience (Edeline & Weinberger, 1993; Weinberger et al., 1993; Edeline et al., 1993; Bakin et al., 1996; Weinberger & Bakin, 1998).

Reorganisation is a natural consequence of the brain's adaptive plasticity. Such a phenomenon can be observed in subjects who have suffered sensory loss. For instance, the brains of blind and deaf subjects can undergo compensatory
mechanisms, whereby the functions of a cortical area normally responsible for processing in the inaccessible sensory modality, are taken over by another sensory modality (Neville et al., 1983; Sadato et al., 1996; Cohen et al., 1997; Kujala et al., 1997; Lessard et al., 1998).

Examples of crossmodal experience-dependent plasticity can also be observed in the adult human brain. For instance, lipreading, an exclusively visual task consisting of watching a speaker's lips during face-to-face conversation, can activate auditory cortex (Calvert et al., 1998; Puce et al., 1998; Weeks et al., 2000). It has been proposed that lipreading improves speech perception by enhancing activity in primary auditory regions, possibly due to integration of the two sensory streams in heteromodal regions around superior temporal sulcus (Calvert et al., 1998). Another example of crossmodal plasticity is primary visual cortex activation during Braille reading and other tactile discrimination tasks (Sadato et al., 1996). Transcranial magnetic stimulation of the occipital cortex in early-onset blind subjects has been shown to induce errors in identification of Braille characters and embossed Roman letters (Cohen et al., 1997). In auditory cortex, plasticity is well illustrated following cochlear implantation in deaf subjects, where long-term cochlear-implant users show a preference switchover from visual to auditory input (Nishimura et al., 2000).

More recently, functional neuroimaging experiments have investigated brain systems involved in making audiovisual associations. These studies show occipital cortex activations triggered by an auditory stimulus (McIntosh et al., 1998, 1999) following crossmodal associative learning. Short-term plasticity in auditory cortex has also been demonstrated by magnetoencephalography (MEG), where a specific frequency band removed from the subjects' acoustic environment for three hours, on three consecutive days, subsequently led to diminished neural representation of that frequency band (Pantev et al., 1999).
The aim of this study was to assess whether plasticity can be demonstrated in the normal human brain during associative learning where learning-related neural responses are indexed by fMRI. Two experiments are described, examining the effect of learning simple audio-visual associations or visuo-auditory associations. From a psychological perspective, in both cases, stimulus presentation in one sensory modality becomes predictive of presentation of a stimulus in a different modality. According to basic principles of associative learning, a contingency, or temporal correlation between two events, should develop, if the probability of a response given a stimulus is greater than the probability of the same outcome given no stimulus or the same stimulus resulting in no response. Generally, if the contingency between a stimulus and a response is positive, learning is illustrated by a negatively-accelerated (exponential) curve (Shanks and Dickinson, 1987).

In the case of this study, stimuli in different sensory modalities are presented consecutively, and associations are expected to form between items which occur in temporal proximity. A contingency between auditory and visual (in this order or the reverse) items is expected to appear over presentations, following a negatively-accelerated exponential curve. This curve can also be used to model events at the neuronal level. A prediction for these experiments is that visual cortex would be activated by an auditory stimulus (which becomes predictive of a visual stimulus during learning of contingent auditory-visual pairings), and that auditory cortex would be activated by a visual stimulus (which becomes predictive of an auditory stimulus during learning visual-auditory pairings).
methods

Subjects
Twenty-four healthy volunteers (7 females) whose ages ranged between 18 to 30 years participated in the study. Twelve of these subjects took part in the audio-visual experiment and the other twelve in the visuo-auditory experiment. All were free of neurological or psychiatric illness. All subjects provided informed consent. This study was approved by the joint Ethics Committee of the National Hospital for Neurology and Neurosurgery and the Institute of Neurology, London.

Experimental materials and conditions
In the experiment subjects were presented with colours and sounds. The colours (red, green, blue, yellow and purple) were projected onto a screen in front of the subject's field of view. The intensity of colours was high, easing recognition of hue. The sounds (distinct abstract noises consisting of mixed frequencies and various manipulated effects, e.g. echo, reverse, fade-in, fade-out) were presented through headphones and adjusted individually for optimal volume levels. In the audio-visual experiment (experiment 1), there were three sounds and two colours. In this case two of the sounds were presented paired to a colour on 50% of trials and alone on the remaining 50%; the third sound was always presented alone. In the visuo-auditory experiment (experiment 2), there were three colours and two sounds. In this latter experiment, two of the colours were presented paired to a sound 50% of the time and alone the other 50%; the third colour was presented individually all of the time. Conditions are illustrated in Fig.1.
Experiment 1: Audio - Visual associations

- Cond 1: sound1-colour1 (S1a-R1v)
- Cond 2: sound2-colour2 (S2a-R2v)
- Cond 3: sound1 (S1a)
- Cond 4: sound3 (S3a)
- Cond 5: colour2 (R2v)

Experiment 2: Visuo - Auditory associations

- Cond 1: colour4-sound4 (S1v-R1a)
- Cond 2: colour5-sound5 (S2v-R2a)
- Cond 3: colour4 (S1v)
- Cond 4: colour3 (S3v)
- Cond 5: sound5 (R2a)

Cognitive task

The behavioural task was to press a button as quickly and accurately as possible every time subjects were presented with R2 (a sound in the case of visuo-auditory associations or a colour for the audio-visual associations). Behavioural performance was measured in terms of reaction times, which were expected to show response benefits when R2 was predictive of S2 (as in cond2: sound2-colour2 and cond5: colour5-sound5 in Fig.1), relative to when it was presented alone (colour2 and sound5 in Fig.1). To measure learning, a comparison was drawn between the change in reaction times over the course of the experiment in response to R2a when it was preceded by S2v versus reaction times to R2a when presented alone. The same measure of learning was adopted in the case of
responses to R2v preceded by S2a versus reaction times to R2v when presented alone.

Both experiments contained two paired conditions. Both paired conditions were expected to produce, with learning, a facilitatory effect for the paired stimulus in response to individual presentation of the first stimulus of the pair, even if the actual contingency of their associations was 50%. One of the paired conditions in each experiment (S2-R2 in both experiments, Fig.1) provided behavioural data by means of a motor response, whereas the other paired condition (S1-R1 in each experiment, Fig.1) provided the neuroimaging data uncontaminated by motor responses.

Changes in brain activation over time during learning were investigated by means of a contrast between the condition where a sound/colour (S1) was presented individually and being predictive of a colour/sound (R1) versus the condition where a sound/colour was always presented alone and thus did not predict another stimulus (S3). Items belonging to the control condition (S3, e.g. sound3 and colour3 in Fig. 1) were always presented unpaired.

**Experimental Procedure**

Two groups of subjects were recruited, one for each study session (audio-visual session and visuo-auditory session). Each group of subjects underwent one study session, while being scanned, where stimuli were presented visually and auditorily. Visual stimuli were presented on a screen located approximately 20 cm. away from the eyes, subtending a visual angle of 7°, and auditory stimuli were presented through fitted headphones and adjusted individually for optimal volume. The presentation duration for all stimuli was 1 s. In the case of paired stimuli, presentation was contiguous and without overlap. The average interval between events, which was randomly jittered (+.375 s, +.750 s, 1.125 s and +1.5 sec.), was 1.875, where paired events involved presentation of two stimuli. There were 26
repetitions of all stimuli which were presented paired and 52 repetitions for those items that were never paired. The order of conditions was randomised, with five stimulus types being randomly presented in blocks of five. The duration of each experiment was approximately 10 min.

**Magnetic Resonance Procedures**

The data were acquired from a 2-Tesla Magnetom-VISION whole-body MRI system (Siemens, Erlangen) equipped with a head volume coil. Multislice T2*-weighted fMRI images were obtained with a gradient echo planar sequence and axial slice orientation (echo time TE=40 sec., repetition time TR=2.935 sec., 64x64x48 voxels). A total of ~200 (201 for subjects participating in the visuo-auditory study and 205 for subjects participating in the audio-visual study) contiguous volume images were acquired per experimental subject. Each echoplanar image comprised 32 2.25mm slices taken every 3mm, positioned to cover the whole of the cerebrum.

Before the first functional scan, one or two scout scans and five dummy scans were acquired to allow for T2 equilibration effects. The functional scans were acquired in an event-related fashion using a trapezoidal sequence. The scanner was synchronised with the jittered onset of events, and the ratio of interscan to interstimulus interval ensured that voxels were sampled at different phases relative to stimulus onset. Structural images were acquired, after the experimental session, sagittally, yielding T1-weighted images with 1x1x1.5 mm. voxel size (matrix size: 256x256x108).

**Data analysis**

The data were analysed using the general linear model, of linear regression on condition-specific waveforms convolved with a haemodynamic response function (Friston *et al.*, 1995a) as implemented in SPM99. More specifically, the time series
were realigned using sinc interpolation adjusting for residual motion-related changes, slice time corrected, spatially normalised (Friston et al., 1995b; Friston et al. 1996) to a standard EPI template (Evans et al., 1992, 1993; Collins et al., 1994), and smoothed spatially with an 8mm. FWHM isotropic Gaussian kernel.

Subsequently a model was specified and estimated using scaling to remove global effects, a high-pass filter providing a basis set of cosine functions with a cut-off period of 61 s (in the audio-visual study) and 75 s (in the visuo-auditory study) to remove low frequency drifts in the BOLD signal (Holmes et al., 1997), and a 4 s low-pass filter.

A group analysis was performed using both fixed and random effects models. The results originating from the fixed effects model were examined to ensure primary effects, particularly visual/auditory activation in response to single visual/auditory stimuli as main effects. For the random effects model, a first-level analysis provided a contrast image for each comparison and each subject. At the second level, these contrast images were the basis of a multi-subject comparison (one-sample t-test). The regressors included in the design matrix were main effects and time parametric modulations (i.e. time-by-condition interactions), modelled as exponential decay functions (i.e. decreases in activation over time for the duration of the whole experiment, i.e. between sessions), for all conditions (i.e. both paired conditions, S1-R1 and S2-R2, S1 unpaired, S3 unpaired, and R2 unpaired) for each subject. There were no null events.

Reported loci of activation which survived voxel-based correction for multiple comparisons were thresholded at p<0.05. A priori hypothesised activations reaching a lower threshold at p<0.001 uncorrected were accepted in order to avoid type II errors.
Behavioural performance

Subjects were debriefed after the experiment on their awareness of the visuo-auditory or audio-visual associations implicit in the stimuli presentations. For each experiment all subjects correctly identified two associations. Behavioural performance was measured as a comparison of reaction times between responding to R2 in the S2-R2 conditions and in the unpaired R2 conditions. In both experiments a significant difference between the paired and the unpaired conditions was found between the median reaction times from each subject as tested by a paired t-test (p=0.01 in the case of the auditory-visual experiment and p<0.01 in the visuo-auditory experiment) (Fig. 2).

FIG. 2. Behavioural results for the visuo-auditory experiment (a) and for the audio-visual experiment (b). Note the difference in reaction times to presentation of R2 acquired over time in both experiments for the paired condition where S2 predicts R2 compared to the unpaired R2 condition where presentation of R2 is not predicted.
Neuroimaging results

1. Main effects of R2 versus S3

Using a group fixed-effects model, the main effect for the comparison between the R2 stimuli versus the S3 was analysed for each experiment. In the audio-visual study, this comparison was between responses to an unpaired colour versus responses to an unpaired sound. This contrast showed activations in visual and motor cortices, both voxel-wise corrected for multiple comparisons, p<0.05 (Table 1). The activations in visual cortex were predicted in response to visual, relative to auditory, stimuli. The motor activations correspond to the motor element of behavioural response present in the R2 condition and not in the S3. In the visuo-auditory study, the same comparison (R2 versus S3) was between an unpaired sound and an unpaired colour. This contrast revealed activations in auditory and motor cortices, both voxel-wise corrected for multiple comparisons, p<0.05 (Table 1).

Table 1. Significant (corrected, p<0.05) activations resulting from the comparison of R2 versus S3 for both types (A-V, V-A) of association.

<table>
<thead>
<tr>
<th>Brain area</th>
<th>Coordinates (x y z)</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A-V</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L intraparietal sulcus/postcentral gyrus</td>
<td>-34 -42 64</td>
<td>6.29</td>
</tr>
<tr>
<td>L fusiform gyrus V4</td>
<td>-32 -70 -18</td>
<td>5.65</td>
</tr>
<tr>
<td>L primary visual cortex V1</td>
<td>-4 -78 -10</td>
<td>5.30</td>
</tr>
<tr>
<td><strong>V-A</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L superior temporal gyrus</td>
<td>-56 -14 -2</td>
<td>6.77</td>
</tr>
<tr>
<td>L transverse temporal gyrus A1</td>
<td>-62 -24 10</td>
<td>6.52</td>
</tr>
<tr>
<td>L precentral/central gyrus</td>
<td>-38 -4 56</td>
<td>5.29</td>
</tr>
<tr>
<td>Mediodorsal thalamic nucleus</td>
<td>6 -16 8</td>
<td>4.81</td>
</tr>
</tbody>
</table>
2. **Visuo-auditory Associations**

The contrast of interest in this experiment was between the unpaired S1v and S3v conditions. Both main effects and time-dependent parametric modulations were analysed.

### 2.1. Main effects: unpaired S1 versus S3

No main effects, i.e. time-independent averaged responses, survived voxel-wise correction for multiple comparisons at threshold p<0.05. At a lower threshold (p<0.001 uncorrected), the comparison of unpaired S1v relative to S3v revealed activations in left inferior (BA46, x,y,z=-16,42,6, Z=4.54) and middle (BA8, x,y,z=-26,20,40, Z=4.48) frontal gyri. The reverse comparison yielded activation foci in right superior temporal gyrus (BA42, x,y,z=68,-24,12, Z=5.04), left primary auditory cortex (BA41, x,y,z=-50,-14,2, Z=4.44) and medial primary visual cortex (BA17, x,y,z=0,-84,-4, Z=4.04).

### 2.2. Time by condition interactions: unpaired S1 versus S3

The *a priori* hypothesis in this study was that a differential response in auditory cortex to a predicted colour would be observed, i.e. a change in activation over time as a consequence of visuo-auditory associative learning. In this respect, predicted regions showed significant time-dependent effects when contrasting visual unpaired S1v versus visual S3v. These activations were localised to left primary auditory cortex (BA41, x,y,z=-44,-34,12; Z=3.85, p<0.001 uncorrected)(Fig. 3 left) and right superior temporal gyrus (BA22, x,y,z=60,-26,4, Z=4.18, p<0.001 uncorrected) (Fig. 3 right). At a lower threshold (p<0.01) both activations were bilateral (x,y,z=44,-28,10, Z=3.46, right primary auditory cortex, BA41; x,y,z=52,-24,4, Z=3.28, left superior temporal gyrus, BA22). Examination of the associated parameter estimates reveals a similar pattern of time-dependent change for both activation foci (Fig. 4).

Fig. 4 represents the “difference” between changes in activation in auditory cortex in response to a visual stimulus that becomes predictive of an associated auditory
stimuli (S1) versus a visual stimulus that does not predict visual stimuli (S3). This means that for the first few trials, when learning is supposed to be maximal, the difference in activation between S1 and S3 is positive, i.e. there is a greater increase in activation in visual cortex in response to the tone that predicts the image relative to the unpredictable tone. This effect is further explained in the discussion.

FIG.3. Coronal sections highlighting activation in A1 (above left) and in STG (above right) resulting from a time parametric modulation contrasting unpaired S1v vs. S3v.
FIG. 4. Graphic plots illustrating the modelled data as a time-by-condition interaction in primary auditory cortex, A1 (above) and in superior temporal gyrus, STG (below). These plots result from the subtraction of parameter estimates belonging to unpaired S1v and the S3v conditions. Note that during the initial learning phase there is a positive difference in activation of these brain areas between the condition where a colour became predictive of a sound relative to the condition where a colour does not become predictive of a sound.

3. Audio-visual Associations
The contrast of interest in this experiment was between the unpaired S1a and S3a auditory conditions. Both main effects and time parametric modulations were analysed.
3.1. Main effects: unpaired S1 versus S3
No effects survived voxel-wise correction for multiple comparison (p<0.05). At a lower threshold (p<0.001 uncorrected), the comparison of unpaired S1a relative to S3a showed activations in left superior temporal cortex (BA21, x,y,z=-52,-48,8, Z=4.53), whereas the reverse comparison revealed no significant activations.

3.2. Time by condition interactions: unpaired S1 versus S3
The a priori hypothesis in this study was that a differential response in visual cortex to a predicted sound would be observed, i.e. a change in activation over time as a consequence of audio-visual associative learning. In this respect, predicted regions of interest showed significant time-dependent effects when contrasting auditory unpaired S1a versus auditory S3a. These regions were localised to left primary visual cortex (V1, BA17, x,y,z=-4,-78,-8; Z=4.59, p<0.001 uncorrected)(Fig 5 left) and left posterior fusiform gyrus (V4, x,y,z=-26,-74,-10, Z=3.74, p<0.001 uncorrected) (Fig. 5 right). At a lower threshold (p<0.01) both these activations were seen to be bilateral (x,y,z=2,-78,-2, Z=3.89; x,y,z=30,-72,-12, Z=3.60). Examination of the associated parameter estimates reveals a similar pattern of change at both activation sites (Fig. 6).

Fig. 6 represents the “difference” between changes in activation in visual cortex in response to a sound that become predictive of associated visual stimuli (S1) versus a sound that does not predict visual stimuli (S3). This means that for roughly the first 10 trials (see X-axis), when learning is supposed to be maximal, the difference in activation between S1 and S3 is positive, i.e. there is a greater increase in activation in visual cortex in response to the tone that predicts the image relative to the unpredictive tone. This difference attenuates over time.
FIG. 5. Coronal sections highlighting activation in V1 (left) and in V4 (right) resulting from a time parametric modulation contrasting unpaired S1a vs. S3a.
FIG. 6. Graphic plots illustrating the modelled data as time by condition interactions in primary visual cortex, V1 (above) and fusiform gyrus, V4 (below). These plots result from subtraction of unpaired S1a and the S3a conditions. Note that during the initial learning phase there is a positive difference in activation of these brain areas between the condition where a sound became predictive of a colour relative to the condition where a sound does not become predictive of a colour.
The traditional view that visual and auditory cortices are tuned to selectively respond to visual and auditory inputs, respectively (Rakic, 1988; Levitt et al., 1997) is challenged by findings in this study. The results illustrate experience-dependent plasticity in response profiles of sensory cortices as a result of associative learning. By means of repeated pairings of audio-visual or visuo-auditory associations, sensory cortices can respond to stimuli in non-specific modalities. Such plasticity has already been shown in animal visual (Bental et al., 1967; Spinelli et al., 1968) and auditory cortices (Edeline & Weinberger, 1993; Weinberger et al., 1993; Edeline et al., 1993; Bakin et al., 1996; Weinberger & Bakin, 1998), in human auditory cortex in lip-readers (Calvert et al., 1998; Puce et al., 1998; Weeks et al., 2000), in human visual cortex in Braille readers (Sadato et al., 1996), and in cochlear-implant users (Nishimura et al., 2000).

Learning visuo-auditory (colour-sound) associations led to auditory cortex being activated during the critical period of association formation (acquisition period) that rendered a colour predictive of a sound, relative to presentations of unpaired colours. Similarly, learning audio-visual (sound-colour) associations led to visual cortex activations during association formation. In other words, in one case a colour evoked auditory activation in left transverse temporal gyrus (primary auditory cortex, A1) and right superior temporal gyrus (association auditory cortex), and, in the other, a sound evoked visual activations in left primary visual cortex (V1) and in left fusiform colour region (V4). Specific neuroanatomical localisation corresponded to that shown in previous studies involving anatomical description of auditory cortex (Pennhune et al., 1996; Rivier & Clarke, 1997; Morris et al., 1998; Belin et al., 1999; Lockwood et al., 1999; Kim et al., 2000; Weeks et al., 2000; Rademacher et al., 2001) and stereotactic maps of visual cortex (DeYeo et al., 1996; Zeki & Marini, 1998; Chawla et al., 1999a, 1999b; Amunts et al., 2000).
It is possible that when two stimuli, in this case sound-colour, are associated due to frequent contiguity in time, presentation of the sound alone can become predictive of the colour, such that at the neuronal level processing the specific sound in auditory cortex leads to activation of visual cortex in expectation of the corresponding visual match or during imagery of the expected stimulus. What is implied here is that auditory and visual cortices communicate. Such communication could take place through mediation of “relay” brain structure, which facilitates their matching by temporal association. That visual cells respond to nonvisual stimuli has been shown before in various species and using different techniques. The activations shown here are not interpreted as a plastic change that takes place during the experiment, i.e. a capacity that cells develop after short exposure, but as an intrinsic ability of visual cells which is expressed in certain situations where stimuli in other sensory modalities lead, e.g. through association to visual stimuli, to activation of visual cortex.

It is notable that the extrastriate focus of activation observed during presentation of auditory items that became predictive of visual stimuli should be localised to V4, an area implicated in colour processing (Livingstone MS & Hubel, 1988; Zeki 1985, 1993; Zeki & Marini, 1998). This finding implies that learning can modulate cortical responses to a high degree of specificity. At a psychological level this result might be explained by the fact that when a sound became predictive of colour, the perception of that specific sound evoked visual imagery of the associated colour, or a general expectation of a colour stimulus.

Activation in V1 was also observed in response to a sound predictive of a colour. It might be that this response is also related to activation of visual imagery. However, there is controversy around the issue of primary visual cortex involvement in imagery (Mitchison, 1996; Roland & Gulyás, 1996). An fMRI study has shown that although V1 and V4 respond during colour perception, these areas do not activate during a task involving colour imagery (Howard et al., 1998). Notably the imagery task used in Howard et al.’s study (mentally comparing colour shades between two
objects of the same hue) was not as simple as the colour imagery which might have been evoked in the present study. In contrast, and in support of a common neural substrate for colour perception and imagery, several neurological studies have indicated that patients with impaired visual percepts, including colour, often have parallel deficits in visual imagery (DeRenzi & Spinnler, 1967; Riddoch & Humphreys, 1987). More specifically, a role for primary visual cortex in imagery has found some support, namely in tasks involving visual recall (Le Bihan et al., 1993) and visual imagery (Kosslyn et al., 1993) in normal subjects. Furthermore, transient application of TMS to BA17 has also been shown to cause impairment in a visual imagery task (Kosslyn et al., 1999).

Specific to this experiment is the experience of colour imagery, which might not occur solely on the basis of primary visual cortex activation but is likely to be coupled with extrastriate contributions. It is possible that V4 inputs into V1 in a feedback fashion. Some evidence in support of this idea comes from an fMRI study, which has shown V4 and V1 activations in response to illusory colour perception as a colour-after effect, known as the McCollough effect (Barnes et al., 1999).

An additional, and not necessarily exclusive, explanation for the presence of activation in primary visual cortex in this study might be related to attentional modulation, where attention is directed to the expected stimulus. It is largely accepted that attention modulates neural activity in extrastriate cortex. For instance, selective attention to colour has been shown to enhance activity in V4 even in the absence of coloured stimuli (Chawla et al., 1999b). However, the extent to which attention operates in striate visual cortex remains controversial (Buckner et al., 1998c), despite the fact that such a finding has already been reported in the context of visual discrimination, where V1 responses were observed to be enhanced to attended stimuli and suppressed when attention was directed elsewhere (Somers et al., 1999). Furthermore, the attentional modulation in this study would have to take place in the absence of a specific visual input. Such a
modulation of activity in the absence of visual stimulation has been reported in extrastriate cortex in the context of directing attention to a peripheral location expecting the onset of a visual stimulus (Kastner et al., 2001).

Modulation of activity was also observed in auditory cortex in response to visual stimuli, where similar mechanisms to that described for visual cortex might be in operation. In simple terms auditory areas may support auditory imagery and/or attentional modulation. Generally, human auditory cortex has not been studied as thoroughly as visual cortex, but it has previously been suggested that primary auditory cortex might be mainly responsible for extracting stimulus features from the environment, whereas secondary regions might be involved in higher-order processes, including internal representation of complex familiar stimuli (Zatorre & Binder, 2000). A previous PET study used a music-based paradigm (pitch judgement to cued words of known tunes) to investigate the comparative neural basis of auditory perception and imagery, and found common activations in superior temporal gyrus (Zatorre et al., 1996). Another study has shown a lateralised effect, where right temporal lobe excision patients performed significantly more poorly on an auditory imagery task (pitch discrimination of visually presented lyrics) than left temporal lobe excision patients or normal controls (Zatorre & Halpern, 1993). Furthermore Penhune et al. (1998) using PET and Rao et al. (1997) using fMRI found right STG activation in response to sequential tapping both with and without the benefit of an auditory cue. In the case of sequential tapping without auditory cueing the pacing is presumably guided by auditory memory or imagery. These studies might support the idea that right auditory cortex might be preferentially involved in imagery processes.

One of the notable features of the present results is the time scale of activations in visual/auditory cortices during crossmodal association formation of auditory/visual stimuli that became predictive of their visual/auditory associates. The time scale was 10 minutes and the main sources of crossmodal effects were expressed during the very early acquisition period. Thereafter, the responses observed in
auditory cortex had an unexpected profile. After the acquisition period, the unpaired visual stimuli evoked relatively greater activation in auditory areas. This result might be a nonspecific effect possibly due to greater attention being focused on the unpaired colour in search of an associated sound. Subjects were not given very explicit instructions as to the structure of the associations, so it is feasible that they were actively searching for associations between stimuli. An alternative explanation is a possible generalisation effect involving an implicit expectation that all colour stimuli might be associated to sounds.

It can also be argued that a visual stimulus alone can suppress responses in auditory cortex and vice versa. In this sense, it is important to take into account that in order to account for deactivations in visual cortex caused by auditory stimuli or deactivations in auditory cortex caused by visual stimuli, one should really have a baseline condition to define “deactivation” against. Unfortunately, this point was not taken into account when designing the experiment and thus the changes in BOLD signal over time can only be interpreted in relation of one condition relative to another. Another study also highlights that the concept of task-specific deactivation from single modality sensory stimulation remains controversial (Laurienti et al., 2002).

Finally, and as a possible methodological caveat, it should be clarified that the use of short ISIs is justified provided that the probability of each stimulus type is random, which in this case it was, and assuming linearity and good model fit. In such cases, the response to one stimulus type cannot be confounded by previous stimuli (Dale & Buckner, 1997; Josephs et al., 1997; Zarahn et al., 1997).

**Conclusion**

These results show, at the behavioural level, exponential response decreases over time in reaction time to stimuli in one sensory modality, which become predictive of stimuli in another modality, relative to unchanged reaction times to unpredictable
stimuli. Associated cortical adaptive responses can also be induced very rapidly, also following an exponential rate of change. Such adaptation was illustrated by auditory cortex activating in response to visual stimuli, which became predictive of certain auditory stimuli, and visual cortex activating in the reverse case. The fact that in a normally developed brain (one where experience has exploited the natural processing resources of different types of neurons) cortical regions which normally respond specifically to one sensory modality, can rapidly learn to respond to stimuli in other senses, exemplifies an underlying capacity for environmental adaptation and challenges previous models of primary sensory cortex function. These findings might have implications in the understanding of specificity of processing in distinct sensory cortices and might open up an area to be explored, namely the role of primary sensory cortices in crossmodal processing and the connectivity underlying such neural operations. Future related studies should also devise a way of directly linking the behavioural condition/task with the experimental conditions underlying the neural operations under investigation.
CHAPTER 7: Crossmodal Transfer

introduction

In everyday life, information is usually processed in more than one sensory modality. For example, visual and auditory attributes of a stimulus often occur simultaneously, thus producing representations accessible through more than one sense. Occurrences such as assigning a voice to a face, or, in the case of objects and animals, and recognising both their defining image and sound are of obvious ecological relevance. Information acquired through multiple sensory channels contributes to formation of multisensory object representation and provides separate sources of access during memory recall.

There is an extensive literature on object recognition pertaining to the visual domain (see reviews; Tanaka, 1993; Beymer & Poggio, 1996; Logothetis & Sheinberg, 1996; Aguirre & Farah, 1998; Mesulam, 1998; Tarr & Buelthoff, 1998; Ullman, 1998; Bovet & Vauclair, 2000). However, relatively little is known regarding crossmodal recognition, namely how a representation of an object encoded in one modality (e.g. visual) can be evoked/retrieved by presentation of the same object in another modality (e.g. auditory). In other words, how the image of an object can cue the retrieval of a sound corresponding to the same object, or vice versa.

Regarding the brain regions involved in such multisensory processing, it has been proposed that information from different sensory modalities undergoes unimodal processing in modality-specific areas, with association areas functioning as convergence zones for information acquired through different sensory streams (Damasio, 1989; Mesulam, 1998). Further theories maintain that activation of polysensory areas (areas where representations formed in one sense are accessible by another) (Pandya & Kuypers, 1969; Jones and Powell, 1970;
Petrides & Iversen, 1976; Stein et al., 1976), or "mediatory" areas, allows different sensory regions to access multisensory representations (Ettlinger and Wilson, 1990).

Neuroanatomical data from monkeys have identified neural structures, defined as sites of convergence of sensory information arising from different sensory modalities, including superior temporal sulcus (STS), premotor cortex, inferior parietal cortex, and anterior cingulate cortex (Jones and Powell, 1970; Seltzer & Pandya, 1978). Impaired crossmodal recognition in monkeys has been reported following lesions to posterior inferotemporal cortex (Sahgal et al., 1975), arcuate cortex (Petrides & Inversen, 1976), and anterior cingulate cortex (Aitken, 1980). Combined lesions of frontal and parieto-occipital polysensory areas or the cortex within the superior temporal sulcus (Streicher & Ettlinger, 1987) produce impairments in crossmodal recognition. Furthermore, aspiration lesions of the monkey amygdala, encroaching directly or indirectly on entorhinal and perirhinal cortices (ventral to the amygdala) (Murray, 1992; Murray et al., 1996; Malkova and Murray, 1996) lead to impairment on crossmodal (tactual to visual) versions of delayed non-matching-to-sample tasks (Murray & Mishkin, 1985). However, it should be acknowledged that subsequent studies involving other modalities, and more specifically visual discrimination based on auditory sensory feedback, showed that selective excitotoxic lesions of the amygdala left monkeys unimpaired on such tasks (Malkova et al., 1997).

There have been few studies involving crossmodal recognition in humans. A notable example is a PET investigation of transfer from the tactile to visual modalities, which evoked activation in the right insula-claustrum (Hadjikhani and Roland, 1998). This region has also been implicated in crossmodal processing, especially in animal studies (Sherk, 1986; Webster et al., 1993). Its extensive connectivity with virtually most of the cerebral cortex, including sensory and prefrontal association areas (Augustine, 1996), makes it a good candidate for a
role in mediating communication between modality-specific areas in the service of multisensory representation formation (Ettlinger & Wilson, 1990).

Another PET study compared intramodal and crossmodal matching in the visual and tactile modalities revealing involvement of claustrum/insular cortex, among other areas, in a crossmodal task (Banati et al., 2000). These two studies are in the visual and tactile modalities and use a crossmodal matching task whereby items are used which have multiple representations; subjects are presented with two objects, either in the tactile, in the visual, or in both modalities; and they have to decide if they are matched or not (if they are the same object). In the present study, the paradigm is more conceptually affine to the idea of crossmodal recognition, where training takes place in one modality and subsequently there is a test phase, where material might be presented in the same modality or in a different modality; during test subjects have to remember whether the information has been presented before and whether it was in the same or in another modality.

The present study investigated crossmodal transfer for material encoded in the visual modality and retrieved in the auditory modality, and vice versa. The material consisted of pictures and sounds representing concrete objects. The principal hypothesis was that multisensory association areas would mediate the transfer of information acquired in one sensory modality at study to another modality at test. Furthermore, specificity of transfer was investigated by contrasting differences in neural networks responsive to visual to auditory transfer and to auditory to visual transfer.
methods

Subjects

11 healthy volunteers (3 females) whose ages ranged between 18 and 40 years participated in the study. All were free of neurological or psychiatric illness. All subjects provided informed consent. This study was approved by the joint Ethics Committee of the National Hospital for Neurology and Neurosurgery and the Institute of Neurology, London.

Cognitive task

Subjects were studied in an event-related paradigm during four separate sessions. Each session involved encoding followed immediately by a test phase. Two of the encoding sessions involved presentation of 20 visual items (line drawings) each, while the other two sessions involved presentation of 20 auditory items (sounds) each. The visual and auditory material was split into separate sessions to avoid saturation and to optimise information retention. Null events (blank screen-no sound) were interspersed within the study material.

During the test phase, half the items (10 items) from each preceding encoding phase were presented in the same modality as during study. The remainder of the study items (10 items) were presented in the alternate modality (see Fig.1). In addition, at each test phase 10 new visual and 10 new auditory items were presented.

The subjects' task at study was to decide whether an item represented a natural or an artificial object (e.g. cow, aeroplane). The subjects' task at test was to recognise each presented item as old (i.e. if they recognised the item as being present during the study phase, regardless of modality of presentation) or new. Subjects performed all encoding and test sessions while being simultaneously scanned.
Experimental materials

The visual stimuli were obtained from the Snodgrass (Snodgrass & Vanderwart, 1980) catalogue and from IMSI MasterClips. The sounds, auditory equivalents of the visual stimuli, e.g. dog barking, telephone ringing, etc., were obtained from IMSI MasterClips and manipulated with SoundEffects to equalise duration of presentation. The study (encoding) blocks consisted of either 20 visual or auditory items and the test (retrieval) blocks contained 20 old items (half visual and half auditory), 20 new items (10 visual and 10 auditory), and 20 null events. The items at test were presented in a pseudo-randomised fashion to avoid patterns but ensuring periodic presentations of the different conditions.

Experimental Procedure

Subjects were MRI scanned throughout the whole study, i.e. study and test blocks. The beginning of a new block was communicated to subjects by visually displaying the words "study" or "test" on the screen. Auditory stimuli were presented through headphones at a comfortable volume regulated individually to each subject. Visual stimuli were presented on a screen located approximately 20 cm. away from the eyes, subtending a visual angle of 7°.

Each stimulus (visual or auditory) was presented for 2s. The inter-stimulus interval, which was randomly jittered (+0TR, +0.125TR, +0.25TR, +0.375TR and +0.5TR) was 3 sec. on average. None of the items were repeated within a study or test session or between different blocks. In total, over the entire experiment, subjects
had 40 visual and 40 auditory items to encode. The duration of the whole experiment was approximately 25 min.

**Magnetic Resonance Procedures**

The data were acquired from a 2-Tesla Magnetom-VISION whole-body MRI system (Siemens, Erlangen) equipped with a head volume coil. Multislice T2* weighted fMRI images were obtained with a gradient echo planar sequence and axial slice orientation (echo time TE=40 sec., repetition time TR=2.935 sec., 64x64x48 voxels). Data were acquired in four sessions. A total of 514 contiguous volume images were acquired. Each echoplanar image comprised 32 2.25mm. slices taken every 3mm., positioned to cover the whole of the cerebrum. Each session comprised an average of 126 volume images.

Structural images were obtained in the same orientation, yielding T1 weighted images with 1x1x1.5 mm. voxel size (matrix size: 256x256x108). An MPRAGE (magnetised prepared rapidly acquired gradient echo) sequence was used, with 108 partitions acquired in the sagittal plane, with a flip angle of 12°, and TE=4ms, TR=9.7ms, TI=600ms. Structural images were acquired at the end of the experiment.

Before the functional scans, one or two scout scans and five dummy scans were acquired to allow for T2 equilibration effects. The functional scans were acquired in an event-related fashion. The scanner was synchronised with the jittered onset of events, and the ratio of interscan to interstimulus interval ensured that voxels were sampled at different phases relative to stimulus onset (a total of one scan/event). This allows the gathering of data to take place at different stages of the haemodynamic response and also it minimises expectation effects.
Data analysis

The data were analysed using the general linear model, specifically linear regression on condition-specific waveforms convolved with a haemodynamic response function (Friston et al., 1995a) as implemented in SPM99. More specifically, the time series were realigned using a sinc interpolation adjusting for residual motion-related changes, slice-time corrected, spatially normalised (Friston et al., 1995b; Friston et al. 1996) to a standard EPI template (Evans et al., 1992, 1993; Collins et al., 1994), and smoothed spatially with an 8mm. FWHM isotropic Gaussian kernel.

Subsequently a model was specified and estimated using scaling to remove global effects, a high-pass filter providing a basis set of cosine functions with a cut-off period of 41 s to remove low frequency drifts in the BOLD signal (Holmes et al., 1997), and a 4 s low-pass filter. A group analysis was performed. Initially a contrast image was created for each comparison and each subject. Subsequently those contrast images were the basis of a multisubject comparison using a random-effects model.

Only scans arising from the test phases were subsequently analysed as conditions of interest. However, all effects (i.e. scans belonging to both study and test phases, including null events) were included in the design matrix. Events were further assigned to different conditions according to whether they were remembered or forgotten at test. This division of conditions applied to both crossmodal (aV and vA) and intramodal (aA and vV) cases, i.e. events corresponding to presentation of items in one modality at study and another at test or those which were presented in the same modality both at study and test. Finally, new visual and auditory items presented during the test phases, were also included in the design matrix. Note that only subsequently remembered events were used to perform the contrasts of interest.
Brain regions are reported as significantly activated due to the comparisons of interest either if they survived voxel or cluster-based correction (at $p<0.05^*$), or at a threshold of $p<0.001$ (uncorrected) for height, or regions with cluster sizes greater than ten voxels.

results

Behavioural data

The mean overall group performance shows significant learning above chance ($p<0.05$) for each study condition (Fig.2), and a significant difference in learning performance between the intramodal and the crossmodal conditions (2-way ANOVA, $p<0.05$). Thus, subjects performed better when the modality of presentation was unchanged between study and test as compared to when the study and test modalities were crossed (two-sample t-test, $p<0.01$). However, another t-test revealed no significant difference ($p>0.1$) in retrieval performance between encoding visually or auditorily, independently of whether there was modality transfer (crossmodal) or not (intramodal). Overall, there was no statistically significant interaction between the two factors of interest: crossmodal/intramodal and visually/auditory encoding.
Fig. 2. Graph showing the percentage of items correctly recognised as old at test. The first two columns show performance levels observed when the material was encoded visually and then retrieved auditorily (red) or visually (blue). The two right-hand columns show performance levels (corresponding to correct hits) when the material was encoded auditorily and subsequently presented visually (red) or auditorily (blue) at test.

**Functional Neuroimaging data**

The results reported are derived from scans resulting from successfully retrieved (recognised) items at test. The main question addressed by the analysis was what brain areas underlie the process of crossmodal transfer, i.e. where are there significant activations when the retrieval modality differs from the encoding modality relative to when encoding and retrieval modalities remain the same (aV versus vV and vA versus aA). In order to address this question several contrasts were performed. Firstly, a contrast compared crossmodal versus intramodal retrieval conditions. It could be argued that this contrast involves a degree of perceptual priming for repeated items (intramodal items). Thus, a further contrast compared crossmodal material versus new items presented only at test, a situation where there is no perceptual priming. The SPMt resulting from this last one-sample t-test was then used to create a mask image (at p<0.05), which was then applied to
the first contrast to isolate the brain regions that were uniquely activated during crossmodal transfer (crossmodal-intramodal masked with crossmodal-new).

1. Crossmodal versus Intramodal

The main effect of crossmodal transfer, i.e. comparing brain areas that activated in response to the items that were successfully retrieved across modality (encoded visually and retrieved auditorily and vice versa) versus those which were successfully retrieved within modality (i.e. encoded and retrieved visually and encoded and retrieved auditorily), yielded activations in left inferior frontal gyrus (BA46-47), left intraparietal sulcus (BA40), and right superior frontal gyrus (BA6) (table 1, Fig 3a). Parameter estimates corresponding to the different conditions are illustrated for the activation focus in left inferior frontal gyrus which reached corrected significance (Fig.3b). Here, a significantly greater contribution to the main crossmodal effect can be observed for the visual retrieval (auditory encoding) condition compared to the other crossmodal condition (auditory retrieval) and both intramodal conditions. Additionally, a simple effects contrast showed that comparing crossmodal versus intramodal visual retrieval (aV vs. vV) yielded activations in intraparietal sulcus bilaterally (x,y,z=-28,-80,34, Z=4.40; x,y,z=32,-72,14, Z=3.57), right intracoccipital sulcus (x,y,z=6,-80,20, Z=3.72), left anterior fusiform gyrus (x,y,z=-40,-64,-18, Z=3.50), and right insula-claustrum (x,y,z=36,8,6, Z=3.70). In contrast, simple effects resulting from comparing crossmodal versus intramodal auditory retrieval (i.e. vA vs. aA) yielded activations in left middle temporal sulcus (x,y,z=-60,-16,-8, Z=3.96), right insula (x,y,z=40,14,16, Z=3.73), left fusiform gyrus (x,y,z=-42,-2,-24, Z=3.60), and left inferior frontal gyrus (x,y,z=-48,38,6, Z=3.55). All reported simple effects were thresholded at p<0.001, uncorrected. Additionally, these conditions were also compared for interactions, where no significant effects were found.
Table 1. Brain activations found when comparing crossmodal versus intramodal retrieval (p<0.05* corrected, p<0.001 uncorrected).

<table>
<thead>
<tr>
<th>Brain area</th>
<th>Coordinates</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>L inferior frontal gyrus (BA47-46)</td>
<td>-44 44 -8</td>
<td>3.76</td>
</tr>
<tr>
<td></td>
<td>-50 40 10</td>
<td>3.70</td>
</tr>
<tr>
<td></td>
<td>-52 28 -4</td>
<td>3.69</td>
</tr>
<tr>
<td>L intraparietal sulcus</td>
<td>-30 -76 34</td>
<td>3.94</td>
</tr>
<tr>
<td>R superior frontal gyrus (BA6)</td>
<td>2 28 58</td>
<td>3.50</td>
</tr>
</tbody>
</table>

FIG 3. (a) Rendered lateral left view with highlighted brain areas, which activated to a greater extent when the modality of retrieval was different from the encoding modality (frontal areas are encircled in yellow and orange, and parietal in blue) relative to when it remained constant; thresholded at p<0.001 for viewing (b) Graphic plot showing the effect size (as parameter estimates -arbitrary units) for the activation which reached cluster-level correction (p<0.05) as a result of contrasting crossmodal (vA and vV) versus intramodal (aA and aV) conditions. Error bars represent standard errors, and the baseline is represented by the null events.

2. Crossmodal versus New

A second contrast (aV versus VNew and vA versus Anew) compared crossmodal recognition (old) (i.e. presented in a modality different to that of encoding) versus new material (new) and showed significant areas of activation in the insula bilaterally, in bilateral inferior frontal gyrus (BA44), bilateral parietal operculum...
(BA40), left cingulate sulcus, and left parietoccipital fissure (table 2, Fig 4). Examination of the parameter estimates for bilateral inferior frontal activations, both of which reached corrected significance at the cluster level, indicated different overall contributions to this effect from various conditions (Fig.5). Specifically, these frontal activations are driven mainly by the crossmodal (auditory retrieval and visual retrieval) conditions. i.e. by vA and aV. Additionally, these conditions were also compared for interactions but no significant effects were found. Subsequently, masking analyses provided a more stringent method of disambiguating the effects of crossmodal transfer (see next section).

Table 2. Brain activations found when comparing crossmodal recognition versus new material (p<0.001).

<table>
<thead>
<tr>
<th>Brain area</th>
<th>Coordinates</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilateral insula</td>
<td>-38 -2 6</td>
<td>4.28</td>
</tr>
<tr>
<td></td>
<td>40 18 2</td>
<td>3.68</td>
</tr>
<tr>
<td>Bilateral inferior frontal gyrus (BA44)</td>
<td>-60 18 16</td>
<td>4.23</td>
</tr>
<tr>
<td></td>
<td>46 12 6</td>
<td>4.21</td>
</tr>
<tr>
<td>L middle frontal gyrus (BA9)</td>
<td>-56 2 40</td>
<td>3.64</td>
</tr>
<tr>
<td>Bilateral parietal operculum (BA40)</td>
<td>44 -26 20</td>
<td>3.86</td>
</tr>
<tr>
<td></td>
<td>-48 -26 20</td>
<td>3.54</td>
</tr>
<tr>
<td>L cingulate sulcus</td>
<td>-8 14 50</td>
<td>3.83</td>
</tr>
<tr>
<td>L parietoccipital fissure</td>
<td>-12 -54 14</td>
<td>3.42</td>
</tr>
</tbody>
</table>
FIG 4. (a) Rendered lateral views highlighting inferior frontal gyrus activations; (b-d) axial sections highlighting bilateral insular (b), left cingulate (c) and left lateral fissure (d) activations; (e) coronal section highlighting bilateral parietal opercular activations; (f) sagittal section highlighting left parietooccipital fissure activation resulting from comparing crossmodal (i.e. retrieval modality different from encoding modality) relative to new (i.e. new items presented at test in the crossmodal modality) conditions; thresholded at $p<0.001$ for viewing.
3. Crossmodal vs. Intramodal masked by Crossmodal vs. New

A more stringent analysis aimed to isolate brain areas activated exclusively during crossmodal transfer, excluding any potential effects due to priming, since it might be argued that reiteration within modality might lead to repetition suppression effects. For this purpose, the two contrasts presented above were masked one by the other, i.e. a mask image (p<0.05) was made from the crossmodal vs. new contrast, which was then used (as exclusive mask) to assess activations unique to
crossmodal relative to intramodal retrieval. This analysis revealed a cluster in left inferior frontal gyrus (BA46) with peak activation in x=-50; y=40; z=10 (Z=3.70) at p<0.001 (uncorrected) (Fig. 6). Note that this area was also activated in the overall crossmodal versus intramodal comparison. Examination of parameter estimates from all conditions shows the contributions to this activation. Note that the greatest effect is driven by the crossmodal conditions relative to intramodal and new conditions (Fig. 7).

FIG 6. Rendered lateral left view highlighting the main activation in inferior frontal gyrus, which resulted from comparing crossmodal versus intramodal retrieval, masked (at p<0.05) by crossmodal retrieval versus new test items conditions.

FIG 7. Graphic plots showing the effect size (as parameter estimates; a.u. = arbitrary units) for activation in inferior frontal gyrus (BA46) found as a result of the masking of Cross versus Intra with Cross versus New. Note the different contributions from the crossmodal (vA and aV), intramodal (aA and vV), and new conditions (AN and VN), with the largest effect being observed for the crossmodal conditions. Error bars represent standard errors, and the baseline is represented by the null events.
The behavioural results showed that recognising items presented in the same sensory modality at study and test was significantly easier, compared to when recognition occurred in a different modality to that at study. There are two possible explanations for the difference in performance between intramodal and crossmodal conditions. One is based on a transfer appropriate processing effect (Morris et al., 1977) or facilitation of retrieval, which requires the same kind of processing employed at encoding. The other interpretation is that of priming effects, i.e. facilitation at recall when study material has been presented repeatedly (see review, Ochsner et al., 1994). This latter possibility was taken into consideration in the analysis of the neuroimaging data. For instance, one of the effects of interest consisted in comparing brain areas activated to crossmodal versus intramodal retrieval. However, this comparison might have been confounded with a priming or familiarity effect at recall for the intramodal conditions where the same object was either seen or heard twice, at encoding and at retrieval. Instead, the data analysis was adapted to circumvent this limitation, namely by comparing crossmodal items versus new items at test and by masking these results with those of the crossmodal versus intramodal contrast.

At the psychological level, the present task is likely to involve initially perceptual processing of the objects presented, regardless of the modality of presentation. Possible associations to the study items might be elicited in the process of encoding. During the test phase, retrieval of specific associations are cued by presentation of study items either in the same modality or in a different modality. The interest of this study is precisely in investigating which neural substrates mediate this crossmodal retrieval process.
Prefrontal cortex - Inferior frontal gyrus

The neuroimaging data show that comparing both crossmodal versus intramodal retrieval and crossmodal retrieval versus new items reveals a focus of significant activation in left inferior frontal gyrus. More precisely, these two contrasts revealed activations in a cluster which extended from the ventral side of BA46 into BA47, and the second contrast alone revealed a peak of activation within BA44. Finally, masking the first contrast with the second revealed a focus of activation in the ventral part of BA46 (see fig. 8). Thus, left inferior frontal gyrus seems to be a key locus in the process of crossmodal retrieval.

In terms of its anatomy, this region receives major polymodal sensory inputs, rendering it a good candidate for convergence of information (Pandya & Kuypers. 1969; Jones & Powell, 1970; Chavis and Pandya, 1976). Furthermore, this ventral prefrontal area, also known as the ventral prefrontal trimodal convergence zone, receives input from second-order sensory association areas (Pandya and Yeterian, 1985). Notably, the human homologue of this area corresponds to the ventral part of Brodmann area 9/46. This region has been implicated in behaviour contingent upon synthesis of information from more than one modality (Passingham, 1972; Passingham & Ettlinger, 1972). Synthesis or convergence of information in more than one modality is required in the present crossmodal transfer task since various sensory representations are used to access the same object, crucially facilitating associative processes involving more than one sensory modality.

Fig. 8 Graphical representation of Brodmann’s areas, lateral view, highlighting the ventral part of BA46.
The findings reported here are consistent with previous functional characterisations of prefrontal cortex in memory. According to this, the mid-dorsolateral frontal region participates in free recall by virtue of its role in on-line monitoring, whereas the ventrolateral frontal cortex is more directly involved in active retrieval mechanisms (Petrides, 1995). In the present study, the retrieval task was more effortful when there was crossmodal transfer than when the test items were presented in the same sensory modality as during study, as verified by behavioural data. Thus, ventrolateral prefrontal cortex (ventral side of BA46) could play a role in matching existing semantic representations of study items with the current (test) presentations, a process which requires significant retrieval search. Similarly, left anterior prefrontal cortex (including BA46) has been reported to show increased activation during demanding retrieval of perceptually detailed information about the studied objects (Ranganath et al., 2000). This latter finding also fits well with the present results in so far as the task in this study placed higher retrieval demands of perceptual information when retrieving across modalities than when retrieving within the same sensory modality.

Ventrolateral prefrontal cortex is also suggested to play a role in active encoding and retrieval of specific information held in visual, auditory and somatosensory association areas, thus allowing for selection, comparison and decision processes regarding information held in short- and long-term memory (Petrides 1994, 1995). Additionally, this region has been linked to associative learning processes (Passingham et al., 2000). It is therefore not surprising that retrieving information in a sensory modality different to the one in which the material was studied, should lead to either formation of multisensory associative representations, or, in this case, retrieval of these associations. It is also noteworthy that damage to the homologue of this region in monkeys leads to impairment on tactile-visual crossmodal matching (Petrides & Iversen, 1976).

Additionally, the material used in this study consists of concrete sounds and images, a factor that ensures access into the semantic system is facilitated. In
other words, retrieving information in a modality different from that in which the material was studied is likely to involve semantic access to multiple representations of the same object. In this regard, left inferior prefrontal cortex has been reported to be part of a semantic executive system that contributes to the online retrieval of semantic information (Buckner et al., 1995; Demb et al., 1995; Démonet et al., 1992; Kapur et al., 1994b; Martin et al., 1995a; Petersen et al., 1989; Petersen et al., 1990), and, more specifically, as a key neural substrate underlying associative semantic processing (Dolan & Fletcher, 1997; Thompson-Schill et al., 1998). This region is also important in semantic processing as demonstrated by depth of encoding manipulations (Kapur et al., 1994a). It is thus feasible that the same prefrontal area is contributing to semantic retrieval when crossmodal transfer is involved, i.e. when associations across different modality representations are activated. Moreover, evidence from previous studies has implicated human inferior frontal areas in semantic, relative to non-semantic, tasks (Ricci et al., 1999) when presentations are in the visual or the auditory modality (Chee et al., 1999), suggesting these areas constitute an amodal system, which allows retrieval of meaningful associations independently of presentation modality.

Another relevant factor in this study was the type of retrieval task presented, i.e. cued recognition. A related task, the ‘remember-know’ procedure (Tulving, 1985), whereby an R judgement indicates distinct episodic recollection, and a K judgements indicates familiarity, has been used in previous neuroimaging experiments. A study based on this procedure found increased activation in left prefrontal cortex, in the same region highlighted by the present study, during recollection relative to new judgements (Henson et al., 1999a). This finding was interpreted in the context of previous reports of left inferior prefrontal activations in relation to source retrieval tasks (Nolde et al., 1998, 1999; Janowsky et al., 1989; Shimamura et al., 1990). This account proposes that this region might play a role in conscious retrieval, possibly involving reflective processes related to source memory. Indeed, in the present study there were clearly increased demands on recollection when identifying objects presented in a different modality at test and
study. This necessarily entails source memory processes, which would explain why activation in this left prefrontal region was greater for the crossmodal retrieval conditions than for the intramodal conditions.

Regarding other foci of activation, areas were found which, together with left prefrontal cortex, might play a role in crossmodal transfer. For instance, the insula activation found when comparing crossmodal old versus crossmodal new fits well with previous findings of insular involvement in crossmodal processing (Hadjikhani and Roland, 1998; Banati et al., 2000). Generally, the strategic location of the insula explains its wide connectivity with the rest of the cortex (see review by Augustine, 1996). Anterior portions of the insula are connected to the frontal lobe, whereas posterior insula is connected to both the parietal and the temporal lobes (Tuere et al., 1999). In this study, this region might be involved in mediating the influx of sensory information from association cortex into prefrontal areas.

The critical finding here is that left prefrontal cortex activates more to crossmodal transfer. This region has previously been found to activate during recognition or cued recall (Henson et al., 1999a). According to Ettlinger & Wilson (1990) crossmodal processing does must not qualitatively differ from intramodal processing. In this sense, it is possible that this region is only mediating the process of recognition but that greater activation signals greater retrieval efforts since behaviourally at least crossmodal retrieval proved to be harder than intramodal retrieval. According to Ettlinger & Wilson (1990), the fact that this region activates for this task does not necessarily imply that it is mediating the convergence of inputs from the training and the test modalities. That this prefrontal region is acting as a polysensory area is not parsimonious with Ettlinger & Wilson (1990) either since they claim that polysensory areas might play a role in crossmodal processing only when the material is unfamiliar, which is not the case in our task.
The present results also highlighted parietal activations, specifically in intraparietal sulcus and in parietal operculum. These are within the classification of post-rolandic or parietotemporal multisensory areas (Seltzer and Pandya, 1980). Detailed neuroanatomical subdivisions of the intraparietal sulcus have been described in nonhuman primates (see General Introduction). These subdivisions comprise specialised multisensory sites for the integration of visual, auditory, somatosensory, and vestibular inputs. However, the level of description of this area in the human brain is much poorer, and MRI results are unlikely to find a match with the more precise monkey subregional classifications. It is, however, notable that an area known to participate in multisensory integrative functions in nonhuman primates is also activated in the crossmodal task of information transfer from visual to auditory modality and vice versa. Indeed, in a recent neuroimaging study, it was suggested that this area plays a role in polymodal motion, i.e. perception of auditory, visual, and tactile moving stimuli (Bremmer et al., 2001).

Conclusion

This study highlights the importance of left ventrolateral prefrontal cortex in the process of crossmodal retrieval. Neuronatomically, this area makes a good candidate for information processing functions which involve multisensory integration, due to its multiple sensory input system. Functionally, this area has been suggested to be involved in active retrieval of verbal, semantic, and perceptually demanding information, as well as in crossmodal cued recognition and associative learning. Thus, the notion supported by previous theories that this area acts as a mnemonic control makes sense in the context of allowing unimodal representations to be matched as belonging to the same object or for separate but associated items to be given a semantic value or be dealt with within a context. This study shows left ventrolateral prefrontal cortex involvement in crossmodal cued recognition where semantic associations established for multisensory representations of concrete objects are retrieved. Future studies in this area should
look into the neural processes taking place specifically during encoding, which lead to crossmodal recognition.
CHAPTER 8: General Discussion

The experiments in this thesis describe a functional neuroanatomy underlying associative learning. The emphasis has been on investigating brain structures involved in the formation of crossmodal associations in the auditory and visual modalities.

Summary of Results

As an introduction to the subject of associative learning, I employed the well-known paired-associate paradigm in the first experiment (chapter 3) to describe, at a general level, the brain systems underlying high-level association processes. It was hypothesised that the results from this study would, at least partially, replicate those of a previous study, which focused on cued recall retrieval (Fletcher et al., 1996). Additionally, the influence on activation patterns of variables, such as imagery and semantic relatedness, contained in the study material during word paired-associate learning, was observed as a relative engagement of different brain systems.

Subjects underwent three study sessions/repetitions of word paired associates during scanning, with learning subsequently tested using a cued recall task. The paired associates were categorised along two dimensions, imageability and semantic relatedness. Behavioural performance scores showed significant learning effects. Encoding imageable (concrete), compared to non-imageable, word pairs was associated with activation in parahippocampal gyrus bilaterally (BA36), right posterior cingulate (BA23/30/31), bilateral visual association cortex (BA19) and right inferior frontal gyrus (BA46). Encoding nonimageable (abstract) words, when compared to imaginable ones, was associated with activation in left inferior frontal
gyrus (BA47), left anterior superior temporal gyrus (BA38), and left superior temporal sulcus (BA21/22). By contrast, comparing semantically related versus unrelated word pairs activated right intraparietal sulcus (BA40), right middle frontal gyri (BA8), right posterior cingulate gyrus (BA31), right precuneus (BA7), right middle temporal gyrus (BA21), and right frontomarginal gyrus (BA10). The reverse comparison, i.e. comparing semantically unrelated versus related paired associate learning, yielded activation of left inferior frontal gyrus (BA45), left posterior inferior temporal gyrus (BA20), left posterior cingulate gyrus (BA23/30), and right caudate nucleus.

These findings illustrate which patterns of brain activation are sensitive to the nature of encoded material, namely the degree of concreteness or abstractness of individual items, and the degree of semantic connection between items. Specifically, different brain systems seem to support encoding mediated by imagery or semantic relations. A network of predominantly visual areas was engaged when association formation was pictorially-based and another network of temporal and frontal areas, generally associated with verbal and semantic functions, was observed when encoded items were not easily imageable (abstract) or when the items in the pair held no obvious connection between each other (semantically unrelated). In this latter case, it was hypothesised that verbal strategies were prioritised over pictorial ones in the process of word-pair encoding.

Note that in this first study, material was presented in one modality, i.e. visually. Subsequently, a different type of associative learning was investigated where a new element was introduced, namely the nature of the associations investigated involved items in different sensory modalities. The second and third studies (chapters 4-5) investigated mechanisms for crossmodal learning in the visual and auditory modalities. In one study (chapter 4) I evaluated time-dependent learning effects in two conditions involving presentation of consistent (repeatedly paired in the same combination) and inconsistent (items presented in various paired combinations) pairs. I also evaluated time-dependent changes for bimodal
(auditory and visual) presentations relative to a condition in which auditory stimuli were repeatedly presented alone.

Using a time-by-condition analysis to compare neural responses to consistent versus inconsistent audiovisual pairs, I found significant time-dependent learning effects in medial parietal and right dorsolateral prefrontal cortices. In contrast, time-dependent effects in response to learning inconsistent versus consistent audiovisual pairs, were seen in left angular gyrus, bilateral anterior cingulate gyrus, and occipital areas bilaterally. A comparison of paired (bimodal) versus unpaired (unimodal) conditions was associated with time-dependent changes in left posterior hippocampus for both consistent and inconsistent pairs. The results provide evidence that associative learning for stimuli presented in different sensory modalities is supported by neural mechanisms similar to those described for other kinds of memory processes. The involvement of left posterior hippocampus in bimodal learning for both consistent and inconsistent pairs supports a putative function for these regions in associative learning independently of sensory modality. However, this study failed to look at crossmodal learning specifically, given that the comparison consisted of paired versus unpaired items.

In the second of these studies (chapter 5) I evaluated time-dependent effects in three conditions involving crossmodal (auditory-visual) and intramodal (auditory-auditory and visual-visual) pairs of stimuli. I again found significant temporal modulation of response in left posterior hippocampus during learning of crossmodal pairs and right anterior hippocampal activation during learning of intramodal pairs. These results show that associative learning for stimuli presented in different sensory modalities, is supported by neural mechanisms similar to those required for association formation between items in the same modality. However, in this case, it is possible that different parts of certain structures, such as hippocampus, process crossmodal or intramodal association preferentially.
The main question addressed in this study was whether hippocampus was specifically involved in crossmodal associative learning, as opposed to merely in general associative processes regardless of presentation modality. The replicated involvement of posterior hippocampus supports a putative function for this region in crossmodal associative learning, particularly in the case of inconsistently paired items, where novelty was a crucial factor. The presence of different degrees of familiarity and novelty in items belonging to the consistent and inconsistent conditions, respectively, make it possible to consider that the differences in hippocampal activation sites are related to these factors and not necessarily to the nature of the items (same or different modality) in the pairs.

The other major finding was the involvement of the insula-claustrum complex in learning associations across modalities, possibly by mediating convergence of inputs from unimodal cortices, allowing items, which would otherwise be processed individually, to be clustered as unified representations. According to the neuroanatomical properties of the insula, it is more likely that its involvement in this paradigm is connected to processing material in various modalities, whereas the hippocampal activations generically, are not likely to be specific to this type of processing, as this area is well known to mediate association formation generically.

The subsequent study presented in Chapter 6 was motivated by an unexpected finding during the data analysis for the study described in chapter 4. In this latter study, it was found that when contrasting the auditory control condition (sounds only, which in the paired condition preceded presentation of images) against the visual control condition (images only, which in the paired condition followed the presentation of sounds), a significant increase in activation developed over time in an area of visual cortex located in right fusiform gyrus. The implication of this finding is that visual cortex can become responsive to stimuli in other sensory modalities, e.g. auditory, if the relevant stimulus is predictive of a visual item. Thus, another study (chapter 6) aimed to investigate time-dependent modulation in neural response when auditory and visual stimuli are paired. This study expanded
the previous finding to explore the modulatory phenomenon not only in visual, but also in auditory cortex, according to whether the predictive stimulus is auditory or visual, respectively.

For this purpose, two experiments were performed, one explored cortical modulation in audio-visual association formation, and the other explored visuo-auditory association formation. Twenty-four healthy volunteers took part in an incidental associative task. Half of the subjects received a task involving audio-visual pairs, and the other half visuo-auditory pairs. They were subsequently debriefed for awareness of the contingencies. The results suggest that visual areas (in striate -V1- and extrastriate -V4- cortices) can respond to an auditory stimulus specifically when subjects are learning that a sound is predictive of a specific colour. I also found that auditory cortex (right superior temporal gyrus -BA21/22- and left transverse temporal gyrus -BA41-A1) can respond to visual stimuli when these are becoming predictive of a sound. These findings highlight the possibility that brain regions, which are traditionally found to respond exclusively to information in one sensory modality can also respond to stimuli in other sensory modalities as a result of associative learning. Moreover, these cortical adaptive responses can be induced very rapidly.

Chapter 7 investigated the transfer effect of encoding and retrieving information in different sensory modalities. All subjects encoded visual and auditory material separately and were subsequently tested either in the same modality of study or in the alternate modality. Behaviourally, learning exceeded chance levels for both crossmodal (encoding and retrieval in different modalities) and intramodal (encoding and retrieval in the same modality) conditions. Neuroimaging results show that certain brain areas are specifically involved in crossmodal retrieval, i.e. in appropriately retrieving information in a different modality to that in which the information was encoded. The main areas found to be involved in this process were inferior frontal gyrus (BA46), intraparietal sulcus (BA40), and insula.
This study describes a neural network of brain areas involved in crossmodal processing, specifically in crossmodal recognition of concrete objects presented in a different modality at test to the modality in which they were studied. It is suggested that sensory association regions provide unimodal representations, which can be accessed in different modality-specific and/or polysensory areas via the insula-claustrum. Information in different sensory modalities is possibly integrated into multisensory representations in parieto-temporal multisensory areas and attributed meaning or meaningful associations in prefrontal cortical areas. Prefrontal cortex is known to play a role in mnemonic control (e.g. Wagner et al, 2001). I think this might be an important component of crossmodal processing whereby multimodal representations might require previously unimodal representations to be matched as belonging to the same object or for separate but associated items to be given a semantic value or be dealt with within a context. This might allow retrieval operations involving monitoring and selection of information from different sources (various unisensory representations, old–new information) for the purpose of appropriate decision making.

This hypothesised mechanism of crossmodal processing might illustrate the adaptive nature of cortical circuits in processing information flexibly and accommodating to the multisensory nature of the world.

Learning, Plasticity and fMRI

Since this thesis revolves around the topic of learning and makes use of fMRI for the purpose of investigating the neural underpinnings of this cognitive process, it is pertinent to discuss how plastic changes that take place during learning can be studied with use of fMRI and the issues involved in such procedures.

One important caveat in the study of learning with neuroimaging techniques is the interpretation of signal change over time. The underlying neural systems involved in the process of learning undergo changes at many different levels, at the
molecular and cellular levels, at the synaptic level, and also at the systems level. The detection of BOLD signal in fMRI restricts the scope of description to the system/s level. It is thus important to bear in mind what is exactly seen when assessing fMRI data, and specifically during dynamic changes due to learning.

Functional MRI does not offer a direct causal relationship between synaptic activity resulting from cognitive processes and the brain activations observed. Instead, fMRI offers correlational information regarding changes in blood oxygenation, which are putatively caused by synaptic activity. However, not only are plastic changes in neural circuitry due to learning in humans fairly unknown at the biophysical level; there are relatively few neuroimaging studies, which have concentrated on describing signal change occurring over time during various learning processes. Some well documented exceptions are studies on motor skill learning (Raichle et al., 1994; Petersen et al., 1998) and effective connectivity (Büchel et al., 1999b; Fletcher et al., 1999). Some work in the field of plasticity has also focused on changes in brain morphology which take place during reorganisation after damage (e.g. Weiller et al., 1995). Additionally, developmental processes are also amenable to the study of plastic changes in brain function over time (e.g. Amunts et al., 1997).

Different strategies are available to tackle the question of signal change over time for the same task. In the case of longitudinal studies, the same subjects can be investigated over several sessions. This approach allows consistency of data originating from the same population, whereby increasing proficiency due to practice might reflect differences in BOLD signal change over time. Another approach is that of cross-sectional observation, more often employed in developmental studies, whereby subjects are selected according to factors such as skill ability over different phases of development and the resulting data might originate from different subjects according to their developmental phase. If subjects were to be followed over a long period during development, such studies could last a very long time.
Such is the case, however, in the study of reorganisation after lesions, where the same subject has to be followed during the process of recovery. In this case, the most notable studies have emerged from stroke patients, whose recovery process is normally very fast. In this regard, it could be argued that effects due to session might be smaller within subjects (longitudinal approach) than between subjects (cross-sectional approach). A relevant study shows what variability can be observed within subjects over different sessions (McGonigle et al., 2000).

There are, however, many factors which could vary between sessions and even within subject, and these are difficult to foresee or control. For instance, during a learning study subjects might experience various levels of anxiety according to their degree of practice in the task or their perceived performance. It is also known that the first time that subjects take part in neuroimaging studies, they might lie restless in the scanner and greater motion artefacts and/or spurious activation due to arousal or increased attention might also be observed, relative to subjects who have previously participated in studies involving performance of a cognitive task while lying in a scanner.

Other possible confounds in fMRI studies involving signal change interacting with time are brain activations related to task difficulty rather than to learning processes. With practice, task difficulty could decrease and thus lead to patterns of activation change, which might be erroneously interpreted as underlying learning. Some confounds can be more problematic than others. Factors likely to be present in all conditions over the duration of the study are naturally less complicating than performance-related confounds, which show a condition-specific pattern.

Conceptually, it is relevant to consider what kind of patterns of activation change can be expected in learning experiments. Increased or decreased signal change over time can be due to various factors. One of these is that the time it takes subjects to perform a given task might decrease with practice. Additionally, the amount of neural recruitment can also be reduced or augmented with experience.
Such changes might reflect increases or decreases in the intensity or the extent of brain activations observed in fMRI experiments. With practice on a learning task, subjects might also become aware of certain aspects which may be unperceived before exposure, and attention might shift to different aspects of the stimuli presented.

It is also possible that certain brain regions might be recruited during the initial stages of learning and that others are specifically engaged at later stages. Thus, it is feasible that the changes in activations over time assigned to learning processes might not be observed in the same regions at different stages of the experiment, simply varying in intensity. Instead, it might be that the observed patterns of activation, i.e. the systems of brain areas engaged, look fairly different at the beginning and the end of the learning experiment. In this sense, a relevant issue to take into account when designing a learning experiment with fMRI is the hypothesised time period of acquisition and the possible relative duration of the various learning phases, i.e. initial, intermediate, and practised phases. These might be characterised by means of reaction times or other behavioural measures, or by observation of activations which are known or expected to engage during early learning, for example. Some experiments might benefit from scanning subjects over different experimental sessions over a period of days, weeks, or months, if the expected consolidation time should have such time spans. Other learning tasks, such as the ones used in the studies described in this thesis, were hypothesised to become well practised over a short period of time and changes in BOLD signal were expected during one single scanning session.

Another aspect which is often overlooked is that fMRI offers no direct correlation between observed brain activations and excitatory or inhibitory synaptic activity. Brain activation as reflected in neuroimaging studies has been mainly associated with glutamatergic activity, which is excitatory in nature. However, this does not imply that activation might not also reflect inhibitory activity.
The interpretation of brain activations during learning is further complicated if we consider that certain activated regions might not be necessary for the process in question, but are active for other reasons. In order to isolate areas which are undoubtedly essential for a cognitive process it is necessary to combine fMRI with other techniques, such as TMS or lesion data, which might highlight the function of isolated regions within the context of a specific task. Certain ways in which other techniques can contribute to a better understanding of the neural underpinnings of learning are explained in more detail in the next section.

A final question is that of contextual effects. In other words, whether the brain responds in the same way regardless of where a task is performed. It is possible that processes, such as learning a skill or associations involving concrete or abstract material, might be under different influences if performed in a scanner or in the real world.

Methodological Issues

At a more general level, certain issues related to the reported studies are worth discussing. From the methodological point of view, all studies comprised in this thesis used the method of fMRI, instead of PET, for greater spatial resolution. Logistically all studies were amenable to the fMRI environment. The equipment used consisted of a screen for visual presentation and headphones for delivery of auditory stimuli, both of which are amenable to the fMRI chamber. The issue of background scanner noise simultaneous with auditory presentations was not considered problematic at the time given the argument that the background noise remains constant throughout studies, thus allowing the assumption that no scanner-related auditory activations would be reflected in the data. However, it is now known that during a functional imaging experiment, the scanner noise induces an auditory response that spans two different temporal scales. First, the scanner noise generated by the acquisition of one slice early in the brain volume may induce activation in an imaging slice which covers the auditory cortex and is
acquired later in the same volume - inter-slice noise interference. Second, scanner noise may induce auditory activation that extends across time to subsequent volumes - inter-volume noise interference. Therefore, by manipulating the timing intervals of the slice and the volume acquisitions in the scanning protocol it is possible to reduce the inter-slice and inter-volume noise interference respectively, and independently, of one another. A study has actually compared two sequences that differed in the inter-volume interval, to show that sparse imaging increases the magnitude of sound evoked activation (Hall et al., 1999). In future, these factors should be carefully taken into account.

From the point of view of study design, the use of event-related fMRI, as opposed to blocked design, was justified on the grounds that for certain learning experiments, e.g. those contained in chapters 4 and 5, items belonging to the same conditions needed to be presented in a randomised order to avoid pattern detection over repetitions of the material. Also, the experiments described in chapter 6 were not well suited for a blocked design since the number of stimuli presented was not sufficient to provide blocks of optimal length, and it was likely that subjects would recognise repetitions of events belonging to the same condition. In the study presented in chapter 7, especially, an event-related design was justified since events had to be classified subsequent to scanning as "remembered" or "forgotten", thus an itemised acquisition of data was needed. Additionally, in all studies a semi-blocked approach was used, whereby items belonging to different conditions were presented in random order for one given repetition of the material, the following repetition containing again items of all conditions in random order, and so on, i.e. A B C D - C B D A - D B A C.

In terms of data analysis methods, the preferred approach taken in all studies, except for the one reported in chapter 4, was that of random effects analysis. The rationale for the random-effects approach employed in most of the studies presented here is that learning and memory processes might be subject to individual differences, thus a reliable tool of generalisation of results to the whole
population was advantageous. It is for example the case, as directly reported from the subjects who took part in the studies, that various mnemonic strategies are used by different individuals in order to encode information, even when the instructions given are highly explicit. It is also the case that stimuli presented can evoke different associations to different participants. Thus, individual differences in brain activation, or inter-subject variance, are accounted for in a random-effects analysis. A fixed-effects analysis, which only accounts for intra-subject variance, can lead to results which are potentially confounded by subjects who might activate more regions above the chosen threshold or with a greater intensity, than others. However, the experiment included in chapter 4 involved a fixed effects analysis of the data, which is prone to subject-based biases. This study was the first one to be carried out and, at the time, random effects data analysis tools were not available.

An aspect common to studies included in this thesis is the fact that they all involved learning processes taking place within a very short space of time. The fact that important changes in brain activation were observed by means of time-by-condition interaction analysis and correlating with significant behavioural performance illustrates a form of short-term plasticity associated with learning, which is expressed in the human adult brain.

The study of time-related changes in brain activity would ideally, however, exploit other complementary techniques, in combination with fMRI. Such a combined approach might reveal further details regarding not only the components of the underlying neural systems but also the temporal correlation of the participating structures, structural connectivity of these systems, functional specificity of brain regions involved, etc.

The neuroimaging modalities suggested for the future advancement of this field include functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET), Electroencephalography (EEG) and Magnetoencephalography (MEG), together with complementary methods like Transcranial Magnetic
Stimulation (TMS) and Diffusion Tensor Imaging (DTI). The suggested combination of methodologies can be implemented in order to study normal neural processes and connectivity, as well as cortical plasticity in learning and reorganisation in various patient populations.

Functional magnetic resonance imaging has been proven to be a reliable tool for mapping brain function. The most common use of this technique has been in characterising different networks of brain areas involved in various perceptual and cognitive functions. In other words, segregation of function, or functional specialisation of certain brain areas, has been the main target underlying neuroanatomical descriptions in fMRI studies. However, specific tools within this neuroimaging modality further allow the investigation of functional integration, i.e. how various brain regions function together to coordinate various stages of a process. These tools are known as effective and functional connectivity, and they aim at finding the influence that one neuronal system exerts over another, and the temporal correlation between neurophysiological events, respectively. Thus, it is very likely that the use of such applications in the study of brain systems involved in learning would provide a deeper understanding regarding how the brain integrates information from different sensory modalities.

By understanding how various neuronal systems interact with each other, it is possible to model a connected network underlying complex functions. In the case of crossmodal learning, it is likely that sensory cortices exert an influence over each other in the process of making associations, for instance between auditory and visual stimuli, as well as other areas, which might play a role in binding these features, in finding semantic connections between items, or in storing them in long-term memory. In future, the study of interregional connectivity could potentially shed light on the issue of multisensory integration, leading to a greater understanding of how certain areas interact to ultimately produce adaptive behaviour.
Another issue of interest in the study of neural processes as dynamic as learning, is the temporal order of participation of brain areas which are involved in different aspects of information processing, manipulation, storage, and retrieval. In the specific case of crossmodal learning, a model is proposed at the end of this section, which suggests a specific sequence of neuronal events in regions involved in primary sensory processing, in mediating convergence of sensory information from different cortices, in linking new and existing items of information, etc.

In this respect, the use of EEG for the purpose of describing the temporal order of activations in a system, and/or to verify hypothesised models of connectivity, might be very informative. This technique alone has the disadvantage of offering very poor spatial resolution (in the centimetre range), relative to that attained with fMRI (in the millimetre range). New advances in this field allow the application of 64 or even 128 electrode arrays, which thus increases the spatial resolution.

Similarly, fMRI, for instance, offers poorer temporal resolution relative to EEG. Therefore an option, which is increasingly proving successful, is that of combining fMRI with EEG, especially in those cases where not only spatial accuracy of description but also temporal order of activations is relevant. Such is the case when attempting to describe the neural processes involved in learning, or more specifically in crossmodal learning. The main difference between these two techniques is that the BOLD fMRI signal results from metabolic processes whereas EEG reflects neural signal. In this respect, a recent study has shown a correlation between fMRI signal and neural signal reflected by local field potentials (Logothetis et al., 2001).

In the context of traditional neuroimaging applications (EEG, MEG, PET and fMRI), a disadvantage lies in the fact that these techniques yield only correlations of brain activity with behaviour. Thus, the observed activations might only reflect unspecific co-activations, which are irrelevant to the task in question. This problem can be partially tackled by a potentially powerful application developed for fMRI,
transcranial magnetic stimulation (TMS), which can be validated using lesion deficit models, either neuropsychological or induced. Neuropsychological models originate from lesion studies which attribute functional specificity to regions according to the deficit they cause in patients. Many of the lesion studies involve the risk of assuming regional functions in the normal brain by inference from functional deficits in specific regions in the lesioned brain. The possible complication in making such assumptions is that cortical reorganisation following lesions could alter functional neuroanatomical patterns in ways that make it difficult to deduce normal function from the damaged site. To overcome this difficulty and corroborate putative functions of various normal brain regions, TMS can be applied to cause a transient interruption of regional functioning in the normal brain, which should emulate the functional deficit in the lesioned brain, thus avoiding the issue of reorganisation.

More specifically, in processes of crossmodal integration, disrupting function in single areas can reveal which regions are necessary for different aspects of the integrative process. For instance, it can be that primary sensory cortices are mainly responsible for analysing single items of sensory information but that additional areas are essential for the process of binding items in an associative manner. In the same vein, it is possible that the insula-claustrum is not essential in the purely perceptual aspect of crossmodal integration but that it plays a fundamental role in creating multisensory representations. However, TMS application requires that brain targets should be easily accessible on the cortical surface, and this would not be the case for structures lying medially, such as the insula. Additionally, disrupting function in the intraoccipital sulcus or in the superior temporal sulcus might lead to difficulty in retrieving multisensory representations.

Another relevant technical caveat in the study of multisensory integration with fMRI is that activation of voxels to multiple modalities does not necessarily imply that the activated area where such voxels are found can be categorised as multisensory. It is possible that the activated voxels contain neurones which process different
unimodal stimuli and that this finer distinction is impossible to make. However, electrophysiological findings have shown that such distinctions can be made. A way to make these distinctions could be by observation of effects typical of multisensory cells, such as response enhancement or depression is observed. So, for example, if the electrophysiological response of a cell is greater for combined sensory modalities than for single modalities, it can be reliably concluded that the cell in question is likely to be multisensory.

Within the main theme of this thesis is the notion that interactions within brain regions, and between brain activations and time, play an important role in learning. Ways of further tackling the question of interactions with time have been suggested. On the other hand, interregional connections, and specifically the structural neuroanatomy underlying such connections, are mediated by axonal fibres, what is typically known as white matter. Estimation of anatomical connections within the living human brain is now possible by measuring the diffusion tensor of water. This is the basis of another MRI application, called diffusion tensor imaging (DTI). This might be particularly interesting when comparing normal and patient populations in search of normal or altered patterns of connectivity. In other words, DTI is a useful tool to investigate the effect of certain lesions in relation to healthy brains, and also to observe anatomical changes in populations where major reorganisation might have taken place, such as in sensory-loss patients.

In summary, the combination of techniques complementary to fMRI (and PET), such as DTI, TMS, and EEG, is aimed at providing further information about brain integrative processes, namely more precise anatomical descriptions, confirmation of functional effects of lesions, and better temporal resolution, respectively.
**Functional Neuroanatomy of Crossmodal Learning**

From the conceptual point of view, this thesis has contributed to an understanding of the neuroanatomy underlying associative learning processes and more specifically crossmodal association formation. These contributions have been in the domains of abstract and concrete association formation (encoding), anticipatory cortical responses to predictive stimuli, and crossmodal transfer mechanisms between encoding and retrieval (recognition) in different sensory modalities. From the neuroanatomical point of view certain structures have been highlighted for their role in these processes.

Sensory information reaches primary sensory cortices during the initial stages of perceptual processing. Activity in these areas can also be modulated by learning, as shown in chapter 6. In this regard, primary visual cortex can respond to auditory stimuli and primary auditory cortex to visual stimuli. This type of functional plasticity illustrates the adaptive potential of the brain. Subsequently, sensory information is further processed in association cortices, where perceptual memories are likely to be stored. Activity in these areas is also amenable to plastic changes during processes of crossmodal association learning whereby a stimulus in one modality becomes predictive of a stimulus in a different sensory modality. This was illustrated in chapter 4 and further explored in chapter 6.

The process of crossmodal association formation was also investigated in chapters 4 and 5, leading to the conclusion that medial temporal lobe is involved in this kind of learning. In these two studies hippocampal activation was evoked by formation of audiovisual associations. The role of the hippocampus in learning has been widely documented. The fact that this structure is also involved in making associations across stimuli in different modalities implies that the hippocampus performs an amodal function. In other words, it detects novel information and creates new associations regardless of type of material. This pattern-matching function is advantageous in an environment where stimuli are likely to be
associated in many different ways, and augments the chances of creating relevant associations rather than discriminating according to specific rules.

Other brain structures involved in other aspects of association formation have also been implicated in memory in other studies. For instance, prefrontal cortex, and in particular the ventral part, plays an important role in assigning meaning to incoming information. This process can be in the form of retrieving existing associations or as active search for associative links or meaning. The fact that this region receives inputs from various sensory association cortices makes it an ideal candidate for integration of information presented in various sensory modalities.

Generally, prefrontal cortex is known to play a crucial role in integrating sensory information temporally in order to produce behaviour, including language. Behavioural outputs are therefore the result of a sequence of neural events which involve the combination of information from various sensory sources. The intricate connectivity of the prefrontal cortex to the rest of the brain, described in the Introduction chapter, explains how such integrative operations are possible. In fact, there is recent evidence to believe that prefrontal cortex indeed supports the production of behaviour based on the formation of relevant crossmodal, and specifically, audio-visual, associations, both in monkeys (Fuster et al., 2000) and in humans (McIntosh et al., 1999; Gonzalo et al., 2000). Such findings only stress the notion that behavioural outputs mediated by prefrontal cortex activity can result from multisensory integration.

However, it is important to highlight that the actual functions performed by this area are likely to be essentially the same in nature independently of modality of presentation. So, for instance, in the context of learning involving word paired associates (chapter 3) prefrontal cortex played a role in association formation. This again emphasises the processing flexibility affordable by neural systems and supports the amodal model of multisensory integration.
More specifically, Fuster’s suggestion of a role for frontal association cortex in the formation of crossmodal associations towards a behavioural goal (Fuster et al., 2000) is highly relevant since it is based on stimuli which do not share physical dimensions. The idea that auditory (tones) and visual (colours) stimuli can be associated without mediation of a communal scale contrasts with the amodal model of multisensory integration.

According to this model, sensory information, regardless of modality, is coded following some continuous scale (number, size, intensity, duration, etc.). It is conceivable that the processing of a scale which can be expressed in more than one sensory modality might rely on temporal synchronisation of neuronal firing in different unimodal areas. This scale would allow stimuli in different modalities to be judged as equivalent or associated. For instance, intensity and pulse might be scales which are handled by common neural mechanisms for different sensory modalities. This model also explains the role of multisensory cells, which respond to a certain scale regardless of sensory modality. Such could be the case of visual areas which are found to respond to auditory stimuli, or vice versa.

However, in this latter case the question remains whether these cells are truly multisensory or whether their function can be ‘expanded’ by plastic mechanisms. More clearly is the case of other brain regions, such as superior colliculus or prefrontal cortex, which have been found to respond to various modalities and their functional role is suggested to be in integrating various inputs to produce adaptive behaviour. However, electrophysiological evidence also shows that the multisensory nature of information processing can be revealed in unimodal cortices. In this regard, some studies have shown that primary sensory cortex can respond to more than one modality (Bental et al., 1968; Maunsell et al., 1989) although traditionally primary areas are considered to be unimodal.

This is possibly a conflicting area which will surely bring changes in the way we think about the brain and its regional functional specialisation. The conflict might
ultimately lie in whether genetic mechanisms predispose various types of cortical cells, located in distinguishable compartments, to process specific types of information, and how flexible these mechanisms are. From an ecological point of view, which would support the idea that the brain is an organ of adaptation to individual and external needs, it might be justifiable to think that cortical cells remain flexible in their functional boundaries. From a more pragmatic physiological point of view, however, it is feasible that optimal brain functioning might be underlined by such specialisation, where tuning and adaptation can take place but only to a certain extent. The limits of this plasticity are yet to be determined.

It seems conceptually plausible that various degrees of genetic determination in relation to function might exist in different brain regions. In this case, areas which process sensory information in its rawest form might have a lesser degree of functional flexibility than areas which are responsible for processing the finer aspects of the sensory input. This explanation implies that a hierarchy of brain systems exists, according to which each area is in charge of a different stage of information processing. Such a hierarchy would subsequently impose a determined amount of plasticity amenable for each participant of a neural system. This amount of plasticity might be regulated differently in the case of normal brains, where information might be processed similarly, relative to brains which have suffered major trauma. In these cases, especially for those where the trauma occurred early in life, a greater flexibility of assigned functions might be available.

This hierarchical model which integrates the idea of information processing and degrees of plasticity available to different brain areas according to their position in the hierarchy suits the view that keeping the first stages of perception as reproducible as possible might be advantageous in order to maintain a reliable version of the world. Whereas, at later stages, allowing information to be combined in multiple forms, might enrich our perception of the world.
There are two views in multisensory processing regarding where and when perception becomes multisensory. A more traditional view argued that unisensory information is temporally processed first and and at a later stage inputs converge elsewhere to produce multisensory perception. More recent findings show that multisensory perception can indeed occur at an early stage and even in cortical regions traditionally considered to be unisensory (Foxe et al., 2002).

Data from this thesis seem to suggest that crossmodal associative learning, with material to be associated in the visual and auditory modalities, does probably undergo a stage of unisensory processing and a later stage of crossmodal processing whereby one visual and one auditory item become associated. In this case the processing stage crucial to crossmodal learning was suggested to occur in the insula-claustrum (chapter 5). In another study (chapter 6), however, it was observed that once crossmodal associations, where one stimulus becomes predictive of another in an alternative modality, have been learnt, presenting the predictive stimulus alone (e.g. a sound) can evoke activation in the alternative cortex (e.g. visual cortex). This case might not provide enough evidence for an early stage of multisensory processing since it does not reflect activations resulting from simultaneous or contiguous presentations of two items in different modalities, but of one item during or after learning. The possibility that multisensory processing takes place at an early stage and in cortex traditionally thought to be unisensory needs further exploration.

Focusing on the relevance of those brain regions which seem to be mainly engaged by processes involving material in more than one sensory modality, it is important to highlight that their underlying neuroanatomy might somehow determine their facilitatory role in multisensory integration. These areas were described in detail in the General Introduction and have been found in some of the studies presented in this thesis.
The insula-claustrum complex, for instance, was observed in the crossmodal associative learning experiment described in chapter 5. The fact that the insula-claustrum was activated only in the crossmodal association study described in chapter 5, where crossmodal associations were appropriately controlled with intramodal associations, and not the one in chapter 4 where crossmodal associations where merely controlled with single unimodal stimuli, suggests that this area might be involved specifically in mediating multisensory integration, since it was concluded that the results from chapter 4 were not conclusive of crossmodal learning. Insula activations were also observed in the context of retrieving information in a modality different to the one in which encoding originally took place (chapter 7). This process involves accessing matching information from memory stores such that the presented object finds its corresponding representation in a different modality, e.g. the image of a cow accessed by presentation of the sound of a cow.

A candidate area for mediation of this accessing process might be the insula-claustrum. In terms of integrating multisensory inputs within the time dimension, it might be that a convergence area provides the temporal structure of neural firing that allows various unimodal representations to be processed as a unified object. In fact, there is already some evidence (Bushara et al., 2001) that activity in insular cortex is related to detection of onset asynchrony between visual and auditory stimuli. In this same study, correlated activity between insula and superior colliculus was also found. Furthermore, response enhancement and depression have also been observed in these two areas during presentation of matched and mismatched visual and auditory stimuli, respectively (Calvert et al., 2000b).

This argument favours the "relay-points" model of multisensory integration. This model posits that stimuli are processed unimodally and independently from one another but remain accessible to other neural populations, "relay points", which are capable of combining these stimuli to produce multisensory integration. Such mediation might be based on strengthened neuronal connections between the two
sensory areas of interest, e.g. visual and auditory association areas, which facilitate the activation of representations encoded in various modalities. It can be that the process of convergence of various inputs or existing memories into a unified representation necessitates a mediatory area. Such an area could for instance be activated when first encoding such correspondences but also during the process of retrieval of existing multisensory representations.

Regarding the processing of temporal patterns in multisensory integration, it is important to highlight that not only insular cortex and superior colliculus, but also prefrontal cortex, have been found to play a role in this function. There is evidence to suggest that prefrontal cortex is involved in integration of sensory information with motor action, especially when time is a crucial component of such integration (Tomita et al., 1999; Wise et al., 1997; Fuster et al., 1985), but also in integrating new crossmodal associations (McIntosh et al., 1999; Fuster et al., 2000; Gonzalo et al., 2000), which might possibly direct subsequent behaviour.

Within the context of a possible hierarchy in information processing, prefrontal cortex would lie at the top, meaning that this region would be the final integrative stage before motor outputs are generated. In other words, according to this hierarchical model prefrontal cortex would be the link between sensory and motor events. In this respect, it is not surprising that various neuronal populations in prefrontal cortex respond to multiple sensory modalities. So, from this point of view, it is also known that prefrontal cortex sends projections back into sensory association cortex, and this might underlie feedback mechanisms.

However, although both insular/clastrum and prefrontal cortices have been found to play a role in integration across time, a question remains regarding the functional difference between these regions in multisensory processing. One possibility is that insular cortex intervenes at an earlier stage of multisensory integration, maybe by facilitating inputs from various sensory modalities to converge, and that prefrontal cortex links various sensory inputs, which have
already been recognised as belonging to the same percept by means of convergence, with existing semantic links or associated representations, or with subsequent behavioural outputs. In other words, the role of insula/claustrum in temporal synchronisation might be restricted to recognising various sensory events as related to the same percept. Whereas the temporal functions of prefrontal cortex might involve links between sensory information present in working memory and items existing in long-term memory and/or motor functions.

Other multisensory structures are located in the junction between the posterior lobes, particularly in the parietotemporal region, e.g. superior temporal sulcus and intraparietal sulcus. Evidence from the studies presented here especially highlights a role for intraparietal sulcus mainly in the process of crossmodal recognition. It is feasible that a region which receives inputs from various sensory association cortices might be involved in integrating memories of objects which are represented/stored in more than one modality. This region would seem to be an important station for perceptual memory and for integrative functions of modality-based features.

Other studies have implicated the intraparietal sulcus, especially in spatial attentional tasks across more than one sensory modality (Eimer et al., 1999; Bushara et al., 1999; Macaluso et al, 2000). More recently, a neuroimaging study suggested that this area might play a role in polymodal motion, which refers to the perception of auditory, visual, and tactile moving stimuli (Bremmer et al., 2001). Such a task involving perceiving motion in different sensory modalities, i.e. as a moving sound, a moving image, and a moving tactile stimulus, might be classified as crossmodal spatial attention.

Response enhancement to matched audio-visual inputs and a sub-additive response to mismatched inputs have also been found in left superior temporal sulcus (Calvert et al., 2000a). This area has been found to be involved especially in speech stimuli. From the neuroanatomical point of view, it is feasible that, because
of its location and connectivity, this area is particularly engaged by audio-visual functions in language (Calvert et al., 1999; Callan et al., 2001; Calvert et al., 2000a; Raij et al., 2001). This might be why the studies in this thesis did not produce superior temporal sulcus activations.

However, other studies have reported a role for this region in functions which might not necessarily be tightly linked to language, such as selective visual and auditory attention, in processing visual motion, mouth movements, in recognising voices, faces and animals (Kawashima et al., 1999; Ahlfors et al., 1999; Lauwers et al., 2000; Puce et al., 1998; Belin et al., 2000; Chao et al., 1999b). Superior temporal sulcus receives inputs from all major sensory association cortices. Thus, within the hierarchy model proposed here, superior temporal sulcus is likely to perform multisensory operations of the highest level. However, this area also receives inputs from prefrontal cortex, which might be suggestive of feedback mechanisms operating during multisensory integration, if we regard prefrontal cortex as the last stage before production of behaviour.

It is important to consider that, although segregation principles are especially emphasised when describing functional neuroanatomy, integration principles are likely to be equally important. Integrative functions are coordinated by neural connectivity underlying various systems of brain regions performing perceptual and cognitive functions. The application of fMRI in this thesis has focused on segregative descriptions, though integrative accounts are an important aspect in the interpretation of these data. However, as previously mentioned, certain tools are now available, which allow a more accurate description of integrated brain functions. Apart from the effective and functional connectivity tools and the recently introduced diffusion tensor imaging applications, electrophysiological techniques can also be of great use in elucidating the temporal order of regional activation. Here I suggest a possible model of connectivity relevant to multisensory integrative functions, which needs appropriate empirical verification (fig. 1).
Fig. 1. Diagram portraying a model of a neural system for multisensory integrative processing.

This model suggests a hierarchy both in functional assignment to various areas within this system and the temporal order of their participation in the process of multisensory integration. At the bottom of the hierarchy and as the starting point of the process are primary sensory areas where normally only unimodal information is processed. Subsequently sensory association areas can be activated if existing perceptual memories are retrieved or finer aspects of the percept are to be analysed. At this point information from various sensory modalities can be associated or integrated, possibly by mediation of a convergence zone, the insula-claustrum complex. In this area, independent unimodal representations can be unified possibly by means of temporal synchrony. Further along, in polysensory areas more complex functions requiring the integration of information in several
modalities might take place. Finally, prefrontal cortex might be in charge of organising multisensory information by assigning meaning to new representations, organising information within the semantic context of existing representations, and/or guiding subsequent relevant motor functions. This model contains feedback loops between the different processing stages allowing flexibility of function, such as the capacity for primary sensory areas to process information in more than one modality, or for top-down mechanisms involving imagery rather than initial direct sensory inputs.

Regarding evidence favouring past suggested accounts of multisensory integration, namely the amodal and the relay-points models, it is appropriate to conclude that both processing systems are possible and might co-exist. It is feasible that certain brain areas are mainly involved in amodal processing whereas other areas with converging sensory inputs might act mainly as relay points for further processing. In fact, several crossmodal tasks have already shown involvement of, both, areas considered purely unimodal and multisensory areas (Graziano & Gross, 1998; Calvert et al., 1999; Macaluso et al., 2000; Bremmer et al., 2001).

I have referred to ‘amodal’ elements in multisensory integration, implying that certain brain mechanisms do not depend on the modality in which information is presented. For instance, the fact that hippocampus is involved in crossmodal learning merely shows that this structure does not discriminate between different modalities. In this sense, the brain might merely be processing intensity and frequency signals from material presented in different flavours (colour spectrum as visual frequencies, sound spectrum as auditory frequencies, etc.). Unisensory cortices might be specialized in processing one of these spectra and this allows us to distinguish say visual from auditory perception. However, in many cases inputs are related to one another and thus, it might not be the case that the senses stop somewhere and an amodal neural code starts. What makes sensory information distinct for each modality might be the combination of amodal properties (e.g.
frequency and intensity, among others) that characterize information in each sense. Unisensory cortices might be carrying out very similar operations but in slightly different combinations. These similarities might be what allows unisensory cortices to switch from one modality to another, or to take over, for compensatory purposes.

My findings show that certain brain regions are amodal, in so far as the type of processing carried out does not discriminate according to which sense information is coded as. For instance, the hippocampal activations reported in chapters 4 and 5, or the prefrontal activations reported in chapter 7. This, however, does not imply that part of the information processing does not involve encoding sensory modality as a marking feature. For example, when retrieving the visual representation corresponding to a sound, prefrontal cortex might indeed process the image and the sound as two separate but associated entities, just like two items in the same modality might be associated. Since an image and a sound are two distinctly separate entities, certain brain areas might be involved in learning that these two percepts originate from the same object and when perceived separately to retrieve that the object has a linking percept in a different sensory modality.

**Future studies**

The study presented in chapter 6 yields evidence to suggest that the modulatory mechanisms observed are category-specific, i.e. that the region of visual cortex that becomes responsive to auditory stimuli will vary according to the type of visual stimulus predicted in the association. For instance, in this study I showed that activations evoked in visual cortex by auditory stimuli which predicted presentation of colour stimuli, were in primary visual cortex as well as in V4; this latter being an area known for its role in colour processing. A future study could test the hypothesis that if a sound were paired to a colour or to a moving object, we would expect the activation in visual cortex to occur in V4 and V5, respectively. Similarly, in the case of the visuo-auditory associations, the area of auditory cortex activated
could vary according to the type of sound predicted from the preceding visual stimulus. This latter case might be harder to argue empirically, since the functional demarcations in auditory cortex have not been as thoroughly explored as those in visual cortex.

In the discussion of chapter 7 a hypothesised model was formulated of brain regions participating in the process of crossmodal transfer. To test the validity of this hypothesis one would require the application of a method which allowed the temporal tracing (i.e. order of information flow) of activations in the different areas of the suggested system. EEG is a possible candidate for this approach, complementing the more accurate spatial localisation of relevant areas with fMRI.

A possibility which was not investigated in this study is whether the efficiency of retrieval and the neuroanatomy underlying it are different if the study material is encoded in both (auditory and visual) modalities simultaneously, as compared to encoding in one modality only. At the behavioural level it seems to be the case that performance improves when different sensory inputs are perceived as belonging to the same object. Such coherence between modalities can theoretically be achieved either by temporal synchrony or by spatial proximity.

Thus, a possible complementary study to the one presented in chapter 7 would consist of a congruence task where objects are presented in two modalities, either in a congruent audio-visual configuration or in a mismatching configuration where the visual and auditory presentations do not correspond to the same representation. In the case of this suggested study, providing congruent pairs would involve the use of meaningful items, i.e. representations which already have a semantic value, or in other words recognisable items retrievable from long-term memory stores. Thus, the focus might lie in investigating neural structures involved in retrieval of congruent crossmodal associations relative to detection of incongruent pairings. The anticipated results might be similar to the ones reported
in chapter 7 on crossmodal transfer, given that the task in this study also involved crossmodal retrieval of congruent associations.

A further related study could facilitate the exploration of the temporal order of information flow in the process of crossmodal retrieval. This would consist of a crossmodal delayed matching task, with presentation of a stimulus in one modality followed by its corresponding representation in another sensory modality.

All studies presented in this thesis were aimed at investigating the healthy human brain, and thus volunteers were selected on this basis. It is, however, important to acknowledge that much can be learnt about crossmodal integration and reorganisation from studying lesioned brains. Such studies would complement our current knowledge of multisensory integration mechanisms. Technically, the advent of MRI applications, such as diffusion tensor imaging (DTI) might facilitate the study of neuroanatomical connectivity in the lesioned, relative to the normal, brain.

The findings reported in chapter 6, namely that learning can produce visual responses to auditory stimuli and auditory responses to visual stimuli, might have implications in the understanding of specificity of processing in distinct sensory cortices and open up a potential conceptual framework for multisensory integrated systems and functional adaptation in brains with damaged sensory areas. One possible way of investigating such adaptive mechanisms might be through application of neuroimaging techniques, which allow observation of neuroanatomical differences in connectivity between the normal and the lesioned brain.

One area in which this approach might be useful is sound localisation in blind subjects. Psychophysical studies have shown that blind subjects possess superior sound discrimination abilities compared to sighted subjects (Lessard et al., 1998). Characterising the neuronal correlate of this phenomenon, namely comparing connectivity patterns in the brains of blind subjects with those of sighted subjects
using fMRI, would involve application of diffusion tensor imaging (DTI) techniques. Based on previous data from sighted subjects, parietal cortex during sound movement analysis would be expected (Griffiths et al., 1998). Additionally, in blind subjects, activation of the occipital cortex, and more precisely visual motion sensitive area MT/V5 (ffytche et al., 2000) in blind persons might prove to reorganise itself to process sound movements. This additional neuronal resource would explain their superior performance in sound discrimination tasks. To assess the functional relevance of these activations as seen in fMRI, parietal and occipital cortices could also be transiently stimulated in a single pulse TMS study in normal subjects.

Also along these lines, a study of crossmodal plasticity is suggested, which involves a patient population and the techniques described above, the phenomenon of interest being sign language in deaf people. It is known that watching sign language by deaf individuals evokes activity in auditory cortex (Nishimura et al., 2000). However, it might be of interest to study the differences in neuroanatomical connectivity between deaf and hearing sign readers.

A related issue that could be further investigated is whether synaesthesia can come about by learning associations between stimuli in different sensory modalities. The finding that a random colour/sound can be associated with a random sound/colour, and that visual/auditory stimuli can evoke activation in auditory/visual cortex might be used as a possible model to further explore the mechanisms mediating synaesthetic experience. This issue, however, might be too complex to be explained simply by theories of association formation. It could be that an imbalance in certain neurotransmitter modulatory system influencing memory is the main underlying cause of synaesthesia. A neuroimaging study highlighting the possible differences in functional neuroanatomy causing the experience in synaesthetes relative to induced synaesthesia in normal subjects might be of interest.
Final words ...

This thesis has provided functional neuroanatomical descriptions of processes which bridge perception and memory in order to elucidate how information is integrated multisensoryly. The modalities investigated here were restricted to vision and audition but further work should be encouraged to investigate other sensory modalities, especially the chemical senses.
References


Bryant PE (1968) Comments on the design of the developmental studies of crossmodal matching and crossmodal transfer. Cortex 4: 127-137.


Chee M W et al. (1997) Auditory and visual word processing studied with fMRI. Neuroimage 5(4), S557.


Goldman-Rakic PS (1998b) Specialised systems for the processing of mnemonic information within the primate frontal cortex In: The prefrontal cortex. Executive and cognitive functions. Oxford University Press.


Hebb DO (1949) The organization of behavior; a neuropsychological theory. NY, Wiley.


270


Mesulam MM, Van Hoesen GW, Pandya DN, Geschwind N (1977) Limbic and sensory connections of the inferior parietal lobule (area PG) in the rhesus monkey:
a study with a new method for horseradish peroxidase histochemistry. Brain Res. 136: 393-414.


280


Nickerson RS (1973) Intersensory facilitation of reaction time: energy summation or preparation enhancement? Psychol. Rev. 80: 489-509.


Paller KA (1990) Recall and stem-completion priming have different electrophysiological correlates and are modified differentially by directed forgetting.


Petersen SE, Fox PT, Snyder AZ, Raichle ME (1990) Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. Science 249, 1041-1044.


Shanks DR and Dickinson (1987) In: The psychology of learning and motivation vol. 21, pp. 229-261. GH Bower (ed.)


Squire LR (1992) Memory and the hippocampus: a synthesis from findings with


Wernicke C (1874) Der aphasische Symtomencomplex, Breslau: Cohn, Weigert.


