1 Healthy and pathological memory: the underlying mechanisms

Section One: The aims of this book

What were you doing immediately before you picked up this book? This question should cause you little difficulty, but there are people who would find it very hard to answer. For these people, known as organic amnesics, life must be experienced as if they were continually waking from a dream. Brain damage has made them very poor at remembering recently experienced events and at learning new information. It also makes them poor at remembering things that were learnt up to many years prior to the brain trauma. Despite such memory impairments, organic amnesics may have normal or superior intelligence. Not all memory deficits caused by brain damage are like organic amnesia, however. Other patients with lesions different from those responsible for organic amnesia show a very rapid loss of spoken information whilst possessing good longer term remembering of most things. For example, such a person might be unable to repeat back more than two spoken digits even with no delay but be able to give the gist of a newspaper article recounted by someone else on the previous day (something well beyond the powers of an organic amnesic). This kind of short-term memory failure is associated with the language disorder known as conduction aphasia and is clearly distinct from the memory deficits seen in organic amnesia, although both are caused by brain damage.

Brain damage is not, however, an essential prerequisite for memory pathology. More popular, at least in novels and films, are lapses of remembering caused by some unbearable emotional strain – the so-called functional amnesias, in which forgetting is thought to be motivated by an unconscious wish not to recall unpleasant memories. In other words, the memories are too painful for the victim to recall. A recent example is provided by the film *The Return of the Soldier*, based on a novel by Rebecca West. In the story, the hero returns from the trenches, apparently intact, except that he can no longer remember anything about his life at home, including the fact that he is married and that his son died in childhood. All that he can remember are the halcyon days of many years before when he was courting his first love. In the dénouement, the shock of being told about his son restores the lost memories. Cases like this are rare but probably authentic. Furthermore, not only do they not require brain damage, but such damage rarely causes so severe a memory loss of this kind. The nearest approximation is seen in senile dementia, but dements have many intellectual deficits not found in functional amnesics.

These examples of organic and functional memory deficits indicate that mem-

ory breaks down in a number of ways. The main aim of this book is to examine the variety of memory deficits that are caused by brain damage in order to gain insight into how the various kinds of memory are mediated by the brains of intact people. In order to do this, it is important, however, to be able to distinguish between organic and functional causes of poor memory. This is not easy to do, partly because such poor memory often has both organic and functional causes. For example, it is not uncommon to find that murderers cannot recall the circumstances of their crimes. These amnesias are sometimes found in individuals who were drunk when they killed their victims, so it is difficult to determine to what extent their forgetfulness is caused by alcohol-induced brain dysfunction at the time of the crime and to what extent by the need to repress painful memories on later occasions. The question of how we can distinguish between functional and organic memory failures is discussed briefly in chapter 2, which is about the assessment of memory problems.

In human beings, brain damage is an adventitious matter, which usually affects several brain regions that control different functions. Thus patients may simultaneously show several kinds of memory failure as well as other cognitive problems. If they do, it becomes hard to know whether the cognitive and memory failures are related and, indeed, whether there is truly one memory deficit or several, dissociable ones that happen to be occurring together. The analysis attempted in this book seeks to identify and characterize those memory disorders that are not made up of further, more selective memory deficits. In other words, it seeks to construct a kind of 'periodic table' of memory disorders and to analyse what underlies each elementary disorder. This goal can be achieved only by gradual approximations. Its fulfilment depends on the accurate description of the functions that have been lost and on the precise location of the brain lesion that causes such losses. It is difficult to be sure that a memory deficit is an elementary one. The ultimate test is whether smaller lesions in the same area where lesions cause the putatively elementary deficit produce merely a less severe problem of the same kind or whether they ever produce a more selective memory deficit. For example, organic amnesia can be caused by lesions in a brain region known as the medial temporal lobes, and involves poor ability to learn new information and poor memory for information acquired pre-traumatically. If it is an elementary memory disorder, it will be impossible to find cases where smaller lesions of the medial temporal lobes ever cause a problem either with new learning in isolation or with pre-traumatic memories in isolation. Such null hypotheses are notoriously hard to establish, but greater confidence in their truth is achieved if patients can be shown to have very circumscribed lesions and if theories of memory derived from studies of intact people suggest that only one memory process has been lost. These checks are, however, fallible because even small brain lesions may compromise several functional systems and because the study of intact people does not provide powerful evidence for the development of memory theories. Such theories are therefore likely to be inadequate and insufficiently articulated. Realistically, then, this book

will aim to identify those memory deficits that are most selective and to specify what brain damage causes them.

If it is known what brain damage causes elementary kinds of memory deficit, where a single function is lost, then it becomes possible to consider how the healthy brain performs that function in intact people. It is too simple to assume without further evidence that the damaged region itself performs the lost function because the activity of apparently intact and normal brain regions may be disturbed by the lesion. Provisional identification of the damaged region as the one that normally performs the lost function may, however, be supported by studies that record electrophysiological and metabolic activity in the brains of intact people whilst they are learning and remembering. For example, recordings of electrophysiological activity made by electrodes placed in the medial temporal lobes of patients during memory tasks are consistent with the idea that this region helps mediate those memory processes disturbed in organic amnesics. Further confidence in the critical role of damaged structures may be gained through a knowledge of their inputs, outputs, and physiology. This knowledge may also provide significant constraints about how the lost function is performed and may ultimately lead to the development of a theory about how neural structures mediate the relevant function. Such theoretical advances are likely to depend heavily on animal studies. These studies not only allow the modelling of human organic amnesia and detailed physiological analysis of the structures damaged in this syndrome, but also may help determine whether there are more specific memory pathologies that have not been found in isolation in brain-damaged people. In this book, discussion of animal models of human memory pathologies and of the physiology and biochemistry of damaged regions will be directed mainly at organic amnesia, but physiological knowledge will be related to the other organic memory deficits where it is relevant.

Organic memory disorders are caused by particular brain regions working poorly either because they have been destroyed (lesioned) or because a biochemical abnormality makes them malfunction. Such disorders are surprisingly common. Apart from selective disorders, such as organic amnesia and the short-term memory deficit, mentioned in the first paragraph, an incomplete list of well-known conditions that are associated with poor memory of various kinds would include normal ageing and pathological ageing, particularly as in Alzheimer's dementia, Parkinson's disease, Huntington's chorea, multiple sclerosis, depression, and schizophrenia. In addition, more exotic kinds of brain damage have been linked with poor memory. For example, severing the fibres that link the two hemispheres of the human brain, an operation that has been performed on certain epileptic patients to relieve their seizures, has been claimed to cause some difficulty with the learning of new information. A subsidiary aim of this book is to discuss such memory disorders briefly to determine whether they are best construed as compounds of the elementary memory deficits that will be tentatively identified or whether they may involve novel kinds of memory impairment, the

identification of which will require more research. The book is not concerned specifically with remediation, although this will be discussed when it fits naturally with a discussion of biochemical or other matters.

In summary, this book will outline the ways in which brain damage in humans causes memory to break down by constructing a 'periodic table' of memory disorders. The nature of the lost psychological functions underlying these elementary memory deficits will be characterized as fully as possible, and an attempt will be made to identify the lesions that cause them. The nature of the inputs, outputs, and physiology of the damaged regions will be discussed, and, where possible, animal models of particular deficits will be considered, in order to develop hypotheses about how the brain mediates the disturbed memory processes. In pursuing these aims, some attention will be paid to less well-specified organic memory problems, such as those associated with ageing and multiple sclerosis, to see whether they can be understood as compounds of the elementary memory disorders.

Although research on human memory pathologies began about 100 years ago, it has mushroomed only in the past 20 years, during which time it has been guided by theoretical ideas drawn from studies of intact people. For this reason, an outline of these ideas is given in the next section. It has already been indicated, however, that research on the memory of intact people may not be a powerful source of ideas about basic memory processes. Neuropsychologists who study organic memory deficits would therefore be unwise to follow too slavishly ideas drawn from this area. The information they discover may indeed be inexplicable in terms of current memory theories and may lead to the emergence of new ideas about memory. But if this is to happen, patients must be studied by researchers who are not blinkered by current theoretical preconceptions – open-mindedness is essential. The section on theoretical ideas drawn from studies of healthy people is followed by one on the ideas about memory that have emerged from physiological studies of memory in animals. The final section sketches the main memory deficits considered to be elementary in this book and introduces the brain regions, damage to which causes these deficits.

Section Two: Theories derived from the study of normal people

Introduction

Research on memory in normal people has been shaped by three theoretical distinctions. The first and oldest hypothesis, made by Plato and Aristotle, divides memory into three stages: registration, storage, and retrieval. In its inchoate form, this theory provides no new information about how human and animal memory works, as any memory system must logically be divisible into these three stages. Modern interest in it lies in the way the stages should be characterized. The nature

of these characterizations determines the acceptability of the other two theories. According to the first of these, memory for rapidly forgotten information is mediated in a basically different way from memory for slowly forgotten information. Somewhat different versions of this theory have been used to explain short- and long-term memory in humans and physiological data with simpler kinds of memory in animals. The third kind of theoretical distinction differentiates among several kinds of long-term memory.

The potential application of these hypotheses to the interpretation of memory pathologies is obvious. As later chapters will show, memory breakdowns have been ascribed to selective failure of registration, storage, or retrieval; to selective failure of short- or long-term memory; and to selective failures of various kinds of long-term memory. To enable the reader to assess these explanations better, the three major kinds of theoretical distinction will be discussed more fully and the types of evidence relevant to them will be considered.

The division of human memory into the stages of registration, storage, and retrieval in itself says nothing about such memory. What matters is the specification of the characteristics of each stage. To do this, the Greeks relied on analogies that compared human memory to the storage of books in a library or of birds in an aviary. In these analogies, what is registered corresponds to a particular book or bird, storage corresponds to keeping the book in the library or the bird in the aviary, and retrieval corresponds to finding the book or bird later. These analogies say next to nothing about the kinds of information processing involved in registration and retrieval or about the way in which information is organized in storage. They also differ from organic memory in one fundamental respect. In human and animal memory, what is registered is not the thing itself but some representation of it. This commonplace distinction between an object and its representation is not apparent in the library and aviary analogies of memory.

What has been learnt about memory's three stages through a century of research into normal human memory?

Encoding and retrieval

One generalization that has emerged to become fashionable in the past decade is the view, developed from the work of Craik and Lockhart (1972), that things are remembered better if they are registered so as to be meaningful to the learner. This view has been supported by many experiments in which subjects are instructed to process verbal or non-verbal stimuli either in terms of meaning or in terms of physical features. For example, a subject might be told either to say which semantic category a word belongs to or to say whether it is written in upper- or lower-case letters. The subjects usually remember the meaningfully processed material better, even if they were not informed during presentation that their memories would be tested later. Originally, Craik and Lockhart argued that meaningfully processed material was better remembered because it was processed

more deeply, and more deeply processed information was forgotten more slowly. Neither of these notions has stood up very well over the past decade. 'Depth' was a concept that rested on the assumption that semantic or meaningful information was extracted at a later stage in the sequence of information processing than were physical features, such as shape or colour. It has been found, however, that subjects can sometimes report semantic details of presented material more quickly than they can report some physical features (for a discussion, see Lloyd, Mayes, Manstead, Meudell, & Wagner, 1984). This has been taken as indicating that semantic features can be registered fairly early in the processing sequence. Similarly, it has been reported that information about semantic features can be forgotten as fast as information about physical features (Nelson & Vining, 1978). More popular now is the view that meaningfully registered or encoded information is remembered better for two reasons: It is encoded so as to be distinct from other information in memory, and the encoded information is better linked to other information in memory. The effect of these encoding advantages is to render such memories easier to find and identify.

The concept of 'distinctiveness' may be intuitively appealing, but it lacks precision. One way to clarify its use is by conceiving of the registration of events in terms of the encoding of a list of features. The distinctiveness of the encoding would then be determined by the number of features uniquely related to the event. This distinctiveness depends on the conditions of retrieval as well as those of learning (see Bransford, 1979). Encoding that is 'transfer-appropriate' leads to the best memory. For example, if subjects are asked about a word's size at retrieval, they will remember best if they encoded the word's physical attributes rather than its meaning. In other words, memory is best when there is correspondence between features encoded and those that are later used as cues during retrieval.

Distinctive and transfer-appropriate kinds of registration not only lead to good recall and recognition, but also are regarded as effortful forms of encoding. They are so regarded because the encoding of meaningful features typically involves processing additional information not immediately apparent in an experienced event. If these kinds of encoding are effortful, then they should be disrupted when subjects perform other cognitive activities at the same time, and effortful encoding should disrupt other cognitive activities. All effortful processes compete with one another in this way and are probably intended behaviours in that, for example, the encoder can indicate what he is aiming to encode. Hasher and Zacks (1979) have contrasted effortful encoding of this kind with automatic encoding of information. In their view, automatic encoding uses minimal attentional capacity and occurs without intentional direction so that other cognitive activities undertaken at the same time are minimally disturbed. Also, automatic encoding is only minimally disrupted by the simultaneous performance of other cognitive operations. Hasher and Zacks have argued that spatiotemporal features of events are automatically encoded and that this ability is fully developed in young children. More polemically, they have claimed that spatiotemporal features can be encoded

only automatically and are not more effectively encoded if effort is applied. In contrast, they have also argued that people learn to encode some semantic features of words automatically. Kellogg (1980) has even shown that faces can be remembered at an above-chance level when little attention is paid to them during learning, which suggests that some facial features can be encoded automatically. No one denies, however, that face memory is much better when effortful registration can occur undisrupted by simultaneous performance of irrelevant cognitive operations. Both effortfully and automatically encoded features are necessary for good memory. It is interesting, therefore, that deficits in long-term memory have been attributed both to selective failures in automatic encoding and to selective failures in effortful encoding. The plausible underlying assumption is that different brain regions mediate the two forms of registration. The assumption is made more plausible by the sharp distinction drawn by Hasher and Zacks between the processes responsible for automatic and effortful encoding.

Appropriate encoding is necessary, but not sufficient, for good subsequent remembering. This is because what is encoded is not necessarily well stored or stored at all. For good subsequent remembering to occur, further storage processes must take place. It is unreliable to decide what information has been encoded by examining the contents of memory after a delay because some of the encoded information may not have been properly stored. If this point is not grasped, there is a danger of blurring the processes of storage with those of registration. The correct way to decide what has been registered is to find out what knowledge the learner can indicate he has at the time of learning so that a minimal memory load is imposed. The subject can indicate this knowledge either explicitly by an appropriate verbal or non-verbal response, such as pointing to an encoded sensory attribute, or implicitly when his behaviour indirectly reflects a recent experience. Implicit encoding can occur without explicit encoding, but explicit encoding always indicates implicit encoding. For example, if a person is shown a picture with a duck hidden in it, then she is usually unable to report seeing the duck. She may, however, be more likely to give 'duck' as a response when asked to give the name of the first farm animal that comes to mind, thus demonstrating that implicit encoding has occurred. If the duck had been identified initially, however, both forms of encoding would have occurred.

One of the discoveries of modern research is that implicitly encoded information has some impact on memory, albeit of a different kind from that found when the information has been explicitly encoded. The rememberer cannot describe the contents of what has been implicitly registered, but can indicate memory by indirect means. For example, Kunst-Wilson and Zajonc (1979) showed subjects irregular polygons for 0.001 sec each, too brief for conscious perception. Although conscious recognition of these shapes did not exceed a chance level, the subjects showed an aesthetic preference for the polygons they had been shown relative to similar novel polygons. Why is the explicit registration of events necessary for their later recognition, but not for the creation of more indirectly indicated memories? The answer is not yet available, but probably relates to the way in

which the information is registered. In the example of the picture with the duck hidden in it, what is the difference between registering the duck's features explicitly or implicitly? The older view is that what is registered is the same, but for unknown reasons, the registered information is explicitly accessible in the former case only. A more recent proposal (Marcel, 1983) is that explicitly registered information is integrated, whereas implicitly registered information is fragmented. This proposal implies that the representation of an implicitly registered complex consists of a number of unintegrated fragments. Explicit registration means that these fragments are 'glued' together in an integrated whole. The idea can be tested because the components of an integrated representation should act as more effective cues for the rest of the representation than components of a fragmented representation. This idea is still primitive, but it may help explain organic amnesia because patients with this disorder, although still conscious, may store fragmented representations, just as if they were encoding implicitly.

In registration, information is processed so that a representation usually, but not necessarily, can be made of some external event. This representation may then be stored with varying degrees of efficiency. Unfortunately, modern research has had little useful to say about different modes of representation that may be used by the brain. It is not even clear what is meant by asking how information is represented. Thus it is not clear whether the same information about an external event can be represented in several different ways or whether different modes of representation mean that different aspects of the event and things related to it are being represented. For example, it is a matter of argument whether representing an event verbally uses a basically different code from the one used in representing it in terms of imagery. If the codes differ, the verbal one may be digital and the imagery one may be analogue, although no one has yet advanced a convincing test of this possibility. One reason for this failure is that representations cannot be directly examined because we can see their effects only after decoding. At present, psychology lacks a theory of how representations are encoded and decoded. Nevertheless, it seems likely that differences in mode of informational representation exist and, if they do, may well be associated with distinct registration, storage, and retrieval processes. If brain lesion A disturbs one kind of memory but not another, and lesion B has the reverse effect, this could be a weak indication that these two kinds of memory use distinct representational codes. Further evidence for this view would be available if the two brain regions were organized anatomically in very different ways.

How information is represented at registration and how it is represented in storage are closely related issues, and probably even less is known about the latter. Questions concerning the properties of storage are very poorly addressed by the study of intact human memory. Indeed, a recent review of human memory completely ignored storage processes and focused entirely on registration and retrieval systems (Horton & Mills, 1984). This is unfortunate for the neuropsychology of memory because some brain lesions may cause storage deficits. For example, it has been suggested that certain organic disorders of language arise partly because

the patient has lost stored information about words. If so, current knowledge derived from studying intact people will not help in the more detailed characterization of such disorders. It would be useful to have knowledge about storage at two levels. The first level is the microscopic one and concerns those physiological changes within and between neurons that underlie memory. Fortunately, research on animal models has led to a partial elucidation of these processes, as will be described in the next section. The second level is the macroscopic one and concerns the way in which information is organized in the brain during storage. Second-level questions are far harder to answer than first-level questions, but can be divided into subquestions, some of which are easier to broach. Thus it is easier to determine where information is stored in the brain and whether different kinds of information are stored in different places. Furthermore, answers to these questions may be needed to resolve the trickier issue of how information is represented in storage. Preliminary attempts to answer the easier questions are discussed in the next section.

It is likely that the acquisition of complex information about events or general knowledge involves the registration and storage of many distinct features that are linked together in a Gestaltic unit. For example, remembering that one talked to two people on a certain London street yesterday involves storing information about the people's features, what was said, what the street was like, and so on, and all these things must be integrated so that they can be retrieved as a single event. Remembering occurs later when, during retrieval, some features of the originally encoded memory reactivate some of or all the other features. A very influential hypothesis about retrieval is the encoding specificity principle (see Tulving, 1984), which states that retrieval cues are effective only in so far as they have been encoded during original learning. When features constituting part of an original memory are re-encoded, there is (or may be) an automatic reactivation of the rest of the memory that does not further involve voluntary search processes. The more cues encoded at retrieval and the more distinctively they characterize the relevant memory, the better will be the remembering. Evidence shows that changes of background context, mode of presentation of information, and mood or physiological state between acquisition and retrieval lead to worse remembering (see Horton & Mills, 1984). This indicates that even those features unrelated to an event's interpretation form part of that event's representation in memory.

In contrast to Tulving, Jones (1979) has distinguished between cues that depend on intrinsic knowledge (encoded during original learning) and cues that depend on extrinsic knowledge (not so encoded). He argues that whereas intrinsic-knowledge cues are associated with rapid, automatic, and unconscious retrieval, extrinsic-knowledge cues are associated with a slower, more conscious, effortful, and inferential retrieval – sometimes called recollection. Recollection tends to occur when memories cannot readily be accessed. It seems to involve the use of extrinsic knowledge to find cues that will be sufficient to activate the automatic retrieval process. For example, in order to remember what one was doing on one's last birthday, it may be necessary to fill in a considerable amount of background

before arriving at cues that will be sufficient to trigger the key information. The interplay of automatic and effortful processes in retrieval resembles that found at registration. The resemblance may be more than coincidence, for it is implicit in much recent theorizing that there is considerable overlap between the processes of registration and retrieval. Effortful registration and retrieval both involve planning activities, so if the capacity to plan is compromised, one would expect both effortful encoding and recollection to be affected. For example, as will be discussed later, lesions to the frontal lobes of the brain may impair memory at least partly for this reason. Furthermore, overlap is also likely because distinctive encodings require the retrieval of semantic information, whereas retrieval requires the encoding of appropriate cues. But there is less reason to believe that the processes involved in more automatic kinds of registration and retrieval are the same.

Recognition and recall

There has been much discussion of and controversy about the processes required for recognition and recall of memories. An early view was that recall involved two processes: a process of generating candidate memories, and a process of identifying the generated items as false or true memories. Recognition was thought to be usually superior to recall because it involved only the identification process, leap-frogging the generation process. If correct, this generate–recognize hypothesis might explain why recall is more affected than recognition in several organic memory disorders. The hypothesis is, however, almost certainly wrong (for discussions, see Horton & Mills, 1984; Tulving, 1984). Neither automatic nor effortful retrieval involves the systematic generation of candidate memories; accessing is more direct than the hypothesis suggests.

Even so, there is some evidence that recognition and recall depend on different kinds of information. For example, Fisher (1979) has shown that only recall is affected by the strength of the links between memory cues and the thing to be remembered, and only recognition is affected by the strength of the links between the thing to be remembered and background memory cues. Such cues are defined in contrast to the information on which attention is primarily focused. It has been argued by Baddeley (1982) that changes in background context, mood, or physiological state during the time between learning and retrieval, which probably do not alter the interpretation of target items, impair recall but have no effect on recognition. These features have been referred to as extrinsic context because it is plausible to argue that they do not influence the interpretation of the target information. Baddeley's claim about the absence of effects on recognition due to changes in extrinsic context seems to contradict Fisher's demonstration that recognition depends on the strength of the links between the thing to be remembered and background memory cues. The claim is open to challenge, however. For example, changes in mode of item presentation do impair recognition (for