The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System, Second Edition

Jeffrey A. Gray
Neil McNaughton

OXFORD UNIVERSITY PRESS

The Neuropsychology of Anxiety

OXFORD PSYCHOLOGY SERIES

Editors

Nicholas J. Mackintosh
Timothy Shallice
Anne Treisman
James L. McGaugh
Daniel Schacter
Lawrence Weiskrantz

1. *The neuropsychology of anxiety: an enquiry into the functions of the septo-hippocampal system (first edition)*
   Jeffrey A. Gray
2. *Elements of episodic memory*
   Endel Tulving
3. *Conditioning and associative learning*
   Nicholas J. Mackintosh
4. *Visual masking: an integrative approach*
   Bruno G. Breitmeyer
5. *The musical mind: the cognitive psychology of music*
   John Sloboda
6. *Elements of psychophysical theory*
   Jean-Claude Falmagne
7. *Animal intelligence*
   Edited by Lawrence Weiskrantz
8. *Response times: their role in inferring elementary mental organization*
   R. Duncan Luce
9. *Mental representations: a dual coding approach*
   Allan Paivio
10. *Memory, imprinting, and the brain*
    Gabriel Horn
11. *Working memory*
    Alan Baddeley
12. *Blindsight: a case study and implications*
    Lawrence Weiskrantz
13. *Profile analysis*
    D. M. Green
14. *Spatial vision*
    R. L. DeValois and K. K. DeValois
15. *The neural and behavioural organization of goal-directed movements*
    Marc Jeannerod
16. *Visual pattern analysers*
    Norma V. Graham
17. *Cognitive foundations of musical pitch analysis*
    C. L. Krumhansl

18. *Perceptual and associative learning*
    G. Hall
19. *Implicit learning and tacit knowledge*
    A. S. Reber
20. *Neuropsychology of communicative behaviour*
    D. Kimura
21. *The frontal lobes and voluntary action*
    R. E. Passingham
22. *Classification and cognition*
    W. Festes
23. *Vowel perception and production*
    B. D. Rosner and J. B. Pickering
24. *Visual stress*
    A. Wilkins
25. *Electrophysiology of mind*
    Edited by M. D. Rugg and M. G. H. Coles
26. *Attention and memory: an integrated framework*
    N. Cowan
27. *The visual brain in action*
    A. D. Milner and M. A. Goodale
28. *Perceptual consequences of cochlear damage*
    B. J. C. Moore
29. *Binocular vision and stereopsis*
    Ian P. Howard and Brian J. Rogers
30. *The measurement of sensation*
    Donald Laming
31. *Conditioned taste aversion*
    Jan Bures, F. Bermudez-Rattoni, and T. Yamamoto
32. *The developing visual brain*
    Janette Atkinson
33. *The neuropsychology of anxiety: an enquiry into the functions of the septo-hippocampal system (second edition)*
    Jeffrey A. Gray and Neil McNaughton

UN COUP DE DÉS

JAMAIS

N'ABOLIRA LE HASARD

Toute Pensée émet un Coup de Dés

Stéphane Mallarmé

———

'The hippocampus—an organ of hesitation and doubt'

P.V. Simonov, 1974

The Neuropsychology of Anxiety

An Enquiry into the Functions of the Septo-Hippocampal System

SECOND EDITION

JEFFREY A. GRAY
Department of Psychology, Institute of Psychiatry, De Crespigny Park, London

and

NEIL McNAUGHTON
Department of Psychology, University of Otago, Dunedin, New Zealand

OXFORD PSYCHOLOGY SERIES
NO. 33

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in
Oxford New York
Auckland Bangkok Buenos Aires
Cape Town Chennai Dar es Salaam Delhi Hong Kong Istanbul
Karachi Kolkata Kuala Lumpur Madrid Melbourne Mexico City Mumbai
Nairobi São Paulo Shanghai Taipei Tokyo Toronto

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First edition 1982
Second edition 2000 (Published in paperback 2003)

A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
(Data available)

ISBN 0 19 852271 1

Typeset by
Florence Production Ltd, Stoodleigh, Devon

Printed in Great Britain

Preface to the paperback edition

In the comparatively short time since the hardback edition of this book went to press in 2001, there have been three developments which readers of this paperback edition might wish to note.

The first lies in the rapidly moving field of research on the different molecular subunits of the $GABA_A$ receptor, of which there are at least 18 different varieties in the central nervous system (Möhler *et al.* 2002). It has long been known that the diverse behavioural and physiological effects of the benzodiazepine class of anxiolytic drugs are mediated in general by action at receptors allosterically coupled to the $GABA_A$ receptor. It is now clear that different classes of these effects are mediated by benzodiazepine binding sites that are coupled to $GABA_A$ receptors of different subunit compositions. In particular, studies of mice with point mutations in specific subunits of the $GABA_A$ receptor appear to rule out as receptors for the specifically anxiolytic effects of the benzodiazepines all subtypes other than those coupled to $GABA_A$ receptors that contain the α_2 and/or the $\alpha 3$ subunit (Möhler *et al.* 2002; Reynolds *et al.* 2001; G. Dawson, pers. comm. 2003). Consistent with the views expressed in this book, the $\alpha 2$ subtype is strongly enriched in the hippocampus, though also in the cerebral cortex; and the α_3 subtype, in the ascending noradrenergic and serotonergic pathways (Möhler *et al.* 2002). The earliest statements of our hypothesis that anxiolytic drug action is mediated by the hippocampal system and its ascending monoaminergic afferents go back several decades (Gray 1970a, 1982), and they were based largely on behavioural data. It is encouraging to see the hypothesis now supported by the latest neurochemical data.

Second, the analysis of quantitative trait loci (QTL; see Chapter 12, section 12.3) related to emotional behaviour has recently been extended from the mouse to the rat, with results that are again encouraging for our model of anxiety. In a study of an F_2 intercross from the selectively inbred Roman High and Roman Low Avoidance rat strains, Ferandez-Teruel *et al.* (2002) gathered a wide variety of behavioural measures from a range of different tests, including the open field, the elevated plus maze, the startle response, fear conditioning and two-way (shuttlebox) active avoidance. One locus, on chromosome 5, was found to influence behaviour across many measures; and to do so in a manner consistent with the effects on the same measures of anxiolytic drugs. Thus, one of the alleles at this QTL was associated with increased shuttlebox avoidance, decreased fear conditioning to both cue and context, increased time spent in the open arms of the elevated plus maze, increased locomotor activity in the open field, and increased rearing behaviour. These effects are all in the direction observed after administration of an anxiolytic drug. Conversely, there was no association between the chromosome 5 QTL and spontaneous activity, the unconditioned startle response or defecation in the open field – all measures that are unaffected by anxiolytic drugs. Correspondence in this way between psychopharmacological profile and genetic architecture provides good support for our overall integrative model of the neuropsychology of anxiety.

Third, despite the difficulties of conducting experimental research on anxiety in human subjects, the recent, almost explosive, expansion of neuroimaging studies is beginning to pay dividends. In one cleverly designed study using functional magnetic resonance imaging (fMRI), Ploghaus *et al.* (2001) measured patterns of brain activity when the subject was presented with a painful heat stimulus under conditions of either high or low anxiety, the higher level being induced by an associated visual stimulus which indicated the possibility that a more painful stimulus might occasionally be delivered. In the high-anxiety condition fMRI activity was augmented in the hippocampal formation, along with correlated activity in a region of the insular cortex specialised for pain perception. This pattern of results is consistent with our hypothesis (see Figure 1.7) that the hippocampal system helps resolve conflict by augmenting negatively affective bias.

The Ploghaus *et al.* study employed normal subjects. A second study, by Furmark *et al.* (2002), investigated patients with social phobia, applying positron emission tomography to study brain activity during a public speaking task. Activity was measured in a no-treatment control group and before and after treatment of two kinds: pharmacotherapy and cognitive behavioural therapy. Both types of treatment were successful with some but not all patients. In treatment-responders, irrespective of type of treatment, improvement was accompanied by a decrease in regional cerebral blood flow in the hippocampus, amygdala and the periamygdaloid, rhinal and parahippocampal cortices. These findings are consistent with the analysis of social phobia offered in this book as reflecting dysfunction in *both* the amygdala *and* the hippocampal formation (section 11.16), in contrast to the pervasive amygdalocentricity of many contemporary approaches to the neuropsychology of anxiety.

In these diverse ways, then, our model of the neuropsychology of anxiety continues to provide a useful framework for understanding accumulating new data.

March 2003 J.A.G.
A.M.

Preface

This second edition of *The neuropsychology of anxiety* provides an updated theory of septo-hippocampal function and an updated theory of anxiety, both of which we believe are superior to competing theories in either domain. The overall theory is summarized in the first chapter. We believe it provides a single account for normal anxiety and generalized anxiety disorder, and for normal memory and clinical amnesia.

The specific details of the theory are important for its correct working, as are precise definitions of the terms involved. Our interpretation of experimental results often runs somewhat counter to conventional views of the data (for example, we do not accept the view that the firing of a hippocampal cell at a particular point in space represents a 'place field' in any normal sense). To appreciate it fully, then, one must read the whole book, including the appendices available on the internet (http://www.oup.co.uk/neuropsychol_anxiety). This is particularly important since we believe that our theory, like its competitors, must be judged in terms of how far it accounts for *all* of the details of the data reviewed, particularly in the appendices. We believe our theory is unique in the degree (admittedly still incomplete) to which it successfully integrates *all* of the data.

The book is also nearly unique in its attempt to integrate data across so many different domains of the literature. This led to a logistical problem with this second edition: it has become very large. For this reason, the basic argument is laid out with only a modest amount of supporting detail in the main text; for those who require the full version of any specific argument, we have relegated what would otherwise be additional chapters of the main text to the appendices. For those reading this book from a purely hippocampal or memorial perspective these probably represent the most critical portions of the book, and they may find it advantageous to start with the hippocampally oriented appendices (4–9) and then read Chapters 7–10. Chapter 8 of the main text deals with the nature of temporal lobe amnesia and attempts to resolve the conflict between the different current views of 'types of memory' that are sensitive or insensitive to hippocampal formation damage. Hopefully our hippocampal analysis will then persuade them to read the rest of the book. For those interested in anxiety, we recommend that they start at the beginning and we hope that they will then find themselves sufficiently impressed by our arguments for the importance of the septo-hippocampal system to read the remainder of the book. The key arguments with respect to the clinical aspects of anxiety are presented in Chapters 11–13.

Despite the advances of the last 20 years, our theory does not yet claim to be complete. So, within certain artificially limited data domains (e.g. the effects of hippocampal lesions on tests, which are supposedly specific to certain types of memory function), aficionados may feel that certain other theories are superior or at least more parsimonious.

However, as will be seen in the various sections of the book, we believe our theory to be at least the equal of the opposition even within each of these specialized domains. Furthermore, we feel that comparative judgements between the theories should be made

over the *whole* range of the available data and here we believe our theory is clearly superior. No other theory of hippocampal function explains (or even tries to explain) the highly specific effects of all known classes of anti-anxiety drugs on the hippocampal formation; and no other theory of anxiety explains why anxiolytic drugs have effects on supposed tests of reference memory function. We would ask our readers, therefore, not to judge the theory solely on its capacity to beat their favoured hypothesis with respect to some small subset of the data covered in this book, but rather to look at the success (or even applicability) of their alternative hypothesis over the whole range of data integrated here.

The core of the theory has withstood the test of time. This is probably the best reason for taking both its psychological and neural aspects seriously. The core of the theory presented in this edition is the same as in the first. Nonetheless, much of what is said in detail in this edition is new—bringing the theory into register with nearly two decades of new data; contrasting it with the new theories which have arisen in that time; and, in particular, rephrasing many of the arguments in an attempt to eliminate the misunderstandings of the past. For example, we try to make clearer that the 'behavioral inhibition system' is neither involved in all inhibition of behaviour nor involved solely in behavioural inhibition—indeed a theoretically fundamental property is its engagement in *active* risk analysis behaviour. This was stated in the first edition, but we have attempted to give it greater emphasis here.

To keep the different aspects of our argument under control the book is divided into three layers. The first is Chapter 1, which contains an overview of the entire argument. The second layer is Chapters 2–13, which expands on the arguments of Chapter 1 at what we hope is sufficient depth for the majority of readers. The third layer is Appendices 1–10 (available via the internet), which provide additional detail of a type which we expect will be required only piecemeal by those wanting the full detail of some specific argument—and so we would not expect these to be read in full by most readers.

Thus, the uniqueness of our view of anxiety derives from analysis of the neuropsychology and neurophysiology of the septo-hippocampal system. Similarly, the uniqueness of our view of the septo-hippocampal system derives from our analysis of the neural, psychological and ethological effects of anxiolytic drugs. Despite these linkages, *neither in the first edition nor here do we equate anxiety with hippocampal function.* Rather, we view the hippocampus as including in its information processing capacities certain operations that are crucial for the maintenance and elaboration of anxiety. In particular, in the present edition, we see anxiety as resulting, in its most fundamental form, from interaction between the septo-hippocampal system and the amygdala.

Emotion and temporal lobe function are arguably the most important areas currently requiring clarification for psychology and neurology. We believe the theory presented in this book successfully reconciles the older 'emotion' view of the temporal lobe with the more recent 'memory' view. Its arguments, therefore, are relevant to both emotion and memory in general as well as to anxiety in particular, and are relevant to temporal lobe function in general as well as the hippocampal formation in particular.

As with our reason for giving the theory credence, the main reason for producing a second edition is that it has withstood the test of time. The core of the theory was presented in a brief article 30 years ago (Gray 1970b). This evolved over a dozen years

into the full-blown theory of the first edition of the book, which was based on similarities between the effects of anxiolytic drugs and hippocampal lesions. Now, nearly 20 years later, in this second edition, we show that these similarities are more remarkable now than they were initially. The original theory was also based on the fact that anxiolytic drugs change hippocampal electrical activity. We have now shown that this is both more general and more firmly established. Similarly, the theory emphasized the role of monoamine systems in hippocampal function, anxiety, and the relation between them. We show that this role is now clearer and more pervasive, and perhaps simpler.

The theory of the first edition attempted to reconcile concepts of anxiety with the facts of hippocampal function and concepts of hippocampal function with a role for that structure in the elaboration of anxiety. The modified theory of the present volume does the same. Luckily, as will be seen, this reconciliation is now, not only more necessary, but also easier.

While the core of the theory remains unchanged, it would be strange if nearly two decades of experimental and theoretical advances did not require some modification of the details and arguments presented in the first edition.

First, recent data have greatly strengthened the crucial original hypothesis that anxiolytic drugs achieve important clinical effects through the septo-hippocampal system. They have also given us a much clearer idea of the neurophysiological basis for these effects. But, while confirming our original emphasis on the hippocampus as being a crucial site of action of anxiolytic drugs within the septohippocampal system, the new data also show that other structures, particularly the entorhinal cortex, cingulate cortex and amygdala, are important for anxiolytic action.

Secondly, many details of the theory have been expanded and adjusted to bring it into line with recent developments in the analysis of clinical anxiety, the neurology of obsessive–compulsive disorder, and the role of the hippocampus and prefrontal cortex in memory. Despite this alignment with current views of memory, our theory is in one sense the antithesis of those which see the hippocampal system as supporting the formation of memories, or which see it as some type of intermediate store (see Eichenbaum, Otto, and Cohen 1994). We offer, as before, a view of the hippocampus as a primarily inhibitory organ. However, in the present edition the precise role of this inhibition in the control of inappropriate memories is more extensively explicated as a special case of the more general functions of the hippocampal formation. We also emphasize the role of the hippocampus in the generation of active behaviour designed to gain the information with which to resolve a conflict.

We have, between us, been studying anxiety, septo-hippocampal function, and the relation between them, for a total of over half a century. This book, inevitably, includes many ideas which we have jointly or separately expressed on previous occasions. It has also borrowed extensively from many others. Despite this, and despite the fact that the core of the theory remains essentially unchanged, this edition presents a new synthesis of work on anxiety and hippocampal function, one that shows more clearly than ever before that the two must be treated as related, though not identical, topics.

May 2000

N.M.
J.A.G.

For Venus, without whom this book could not have been written; and for Julie, without whom it could not have been rewritten.

Acknowledgements

Jeffrey A. Gray

I am deeply grateful to Professor Jean Scherrer for the kindness and hospitality he extended to me during my stay in his laboratory in the Faculté de Médicine, Université de Paris VI, in 1979–80, when the first edition of this book was written. My thanks are also due to the Nuffield Foundation for the award of a Social Science Research Fellowship which, in 1975–6, allowed me to start working on the book, and for help in the preparation of the manuscript. I am indebted to all of my colleagues and students at Oxford for innumerable ideas and points of information which have influenced my thinking over the years. Particular thanks are due to Dr Marianne Fillenz, for educating me in the mysteries of the noradrenergic neuron; to Professor Nicholas Rawlins, for clarifying many points in hippocampal physiology; to Mrs S. Digby-Firth and Mrs M. Penning-Rowsell for their care in preparing the manuscript; and to my sons, Ramin and Babak, for editorial assistance. In regard to this second edition, I have received great help from Finola Curtin.

Neil McNaughton

I am grateful to all my students who, over the years I have been preparing this second edition, have not only put up with its distraction of me from their work but have helped me with comments, ideas, and friendship. I must also thank the University of Otago and my head of Department Professor Geoffrey White, for their permitting me six months of prolonged research time without which this edition would have appeared many years later than it has. Len Jarrard is owed special thanks for providing Jeffrey and myself with a haven at Washington and Lee University where, for several days half way round the world from each of our places of work, we could meet and thrash out the final details of the present edition.

In producing this second edition, I owe a particular intellectual debt to Robert and Caroline Blanchard, Fred Graeff, Ray Kesner, Ian Kirk and Olga Vinogradova. Many of the modifications to the original theory arose out of discussions with them, making it difficult to acknowledge their precise contribution in the text. Others who have discussed their ideas with me, and who I hope will forgive me for what I have done with those ideas, are Carol Barnes, Rob Berman, Brian Bland, Michael Davis, Joe Le Doux, Joaquin Fuster, Charlie Gross, Len Jarrard, Bruce McNaughton, Sheri Mizumori, Tony Phillips, Bob Sainsbury, Bob Vertes, and Tony Wright. I have tried to acknowledge their contributions in the text but in case I have failed to fully acknowledge my borrowings, I thank them here.

I am also most grateful for those who provided detailed comment on draft chapters: Cliff Abraham, Carol Barnes, Rob Berman, Stuart Checkley, Fred Graeff, Ray Kesner,

Ian Kirk, Brian Young, and Sidney Wiener. Here, particular thanks must go to Carol Barnes and Sidney Wiener for their extensive criticism of my analysis of single cell recordings, which has resulted in major changes—and, I hope they will think, improvements.

Moana Theodore and Marlane Bronstring gave me extensive help with the references. Very many others, far too numerous to mention, have kindly sent us reprints and preprints of their work over the last few years, we thank them for their help here and have acknowledged their work by citation in the text. Joe Ledoux's courier delivery of his book *The emotional brain* to us was particularly appreciated, as was Bob Vertes' hand delivery of a draft of his major review with Bernat Kocsis of *Brainstem–diencephalo–septo-hippocampal systems controlling the theta rhythm of the hippocampus*. Wendy Suzuki kindly supplied electronic copies of a number of complex figures on hippocampal connections and James Freund supplied the original artwork for Figure 9.6.

We are also very grateful to those who responded heroically to a last minute request for preprints and papers which I wished to cite directly but lacked a copy: Robert Adamec, John Aggleton, Efrain Azmitia, David Barlow, Bob and Caroline Blanchard, Brian Bland, Randy Commissaris, John Dalrymple-Alford, Phillip Best, Elena Brazhnik, Sam Deadwyler, William Deakin, Howard Eichenbaum, Don Fowles, Steven Fox, Tamas Freund, Sandra File, Mike Gabriel, Patricia Goldman-Rakic, Donald Klein, George Koob, Serge Laroche, Joe Ledoux, Isaac Marks, Paul Monmaur, Richard Morris, Elizabeth Murray, Lynn Nadel, Tim Otto, Michael Petrides, Tony Phillips, Trevor Robbins, Edmund Rolls, Bob Sainsbury, Larry Squire, Dallas Treit, and Case Vanderwolf.

The updating of this second edition over the last few years has been greatly aided by the immensely supportive environment of the Department of Psychology at the University of Otago and by the sheer number of laboratories in the Department working directly on the hippocampus, memory, and plasticity, each from their own perspective. Thanks for many unacknowledged inputs go to Cliff Abraham (LTP/LTD), David Bilkey (theta/memory), Mike Colombo (single cells/memory), Cynthia Darlington, and Paul Smith (pituitary adrenal hormones and plasticity), Bob Knight (amnesic syndromes), Brian Young (single cells), and Geoff White (signal detection models of memory).

Contents

1 Overview

This book presents a theory of the neuropsychology of anxiety entwined with a theory of the functions of the septo-hippocampal system. This first chapter provides an overview of our entire argument. The details are then progressively deepened throughout the printed text, with still further amplification in the appendices, available electronically at http://www.oup.co.uk/neuropsych_anxiety. In summary, we hold that the septo-hippocampal system acts to detect conflict between concurrently available goals; that it resolves this conflict by increasing the weight given to affectively negative information by goal-processing areas; and that this increase in the weight of affectively negative information produces increases in anxiety in some tasks while reducing the effects of interference on memory in others.

To many people encountering our theory for the first time this juxtaposition of anxiety with the septo-hippocampal system will appear odd, if not heterodox. Our original heresy, in the first edition in 1982, lay principally in holding that the septo-hippocampal system has an important role to play in anxiety when orthodoxy held it to be important for memory. This original heresy remains. But, to it, this second edition adds a new one: for orthodoxy now holds that the key brain structure underlying anxiety is not the septo-hippocampal system, but the amygdala. So, we seem to have got it wrong both ways: our structure of interest, the septo-hippocampal system, is generally held to do something quite different than control anxiety; and our condition of interest, anxiety, is generally held to depend on a quite different structure than the septo-hippocampal system.

We seem, then, to be stubborn, indeed perverse. But it is a stubbornness justified by nearly two decades of data that at many points confirm the 1982 version of the theory. Conversely, the data which supported the distinctive cornerstones of our original theory are unaccounted for by other theories, even though they have become more rather than less well established over these years. In particular:

1. What are now known as the 'classical' anxiolytic drugs (benzodiazepines, barbiturates, etc.) were held in 1982 to be similar to septo-hippocampal lesions in their effects on behaviour. We now know that this parallel holds also for the novel anxiolytics. Unlike the classical drugs, these do not act via synapses at which the endogenous transmitter is γ-aminobutyric acid (GABA); and they do not share any of the side-effects (anticonvulsant, sedative, hypnotic, and addictive) of the classical anxiolytics. We now know also that the parallel with septo-hippocampal lesion effects extends, for both classical and novel anxiolytics, to certain crucial 'hippocampal' tests of memory function.

2. The classical anti-anxiety drugs were held in 1982 to produce specific impairments in the control of 'theta activity'. This is a characteristic phasic pattern of firing of cells in the hippocampus and related structures. It is controlled by pacemaker cells in the medial septal area and, because many cells fire synchronously, can give rise to an easily recorded gross extracellular 'theta rhythm'. We now know: (a) that similar impairments in the control of the theta rhythm are produced by all clinically effective anti-anxiety

drugs, including the novel anxiolytics and anxiolytic antidepressants such as imipramine; (b) that this is so far true *only* for drugs known to be clinically effective in treating generalized anxiety disorder; and (c) that these impairments in hippocampal electrophysiology—when produced by manipulations confined to the specific brain nuclei controlling them—produce behavioural effects equivalent in many respects to systemic administration of anxiolytics. The effects of these compounds on hippocampal electrophysiology can therefore account, in principle, for the bulk of the critical anxiolytic-induced changes in behaviour shared by septo-hippocampal lesions.

What of our perversity? Since 1982 there has been an evolution of our theory, and, in addition, of other theories of the septo-hippocampal system, theories of the neural basis of anxiety, and theories of the psychological nature of generalized anxiety. This evolution has moved them all towards a common centre, reducing (though not eliminating) our heterodoxy. Our view differs from each of the others, most noticeably in attempting to combine them all. But, in several cases, it differs only semantically. So, we shall try to persuade you that the current orthodoxies are, in critical respects, incomplete; and that in this book we have captured the common centre towards which all of them have been converging.

We shall argue that the role of the septo-hippocampal system in the control of memory is to resolve conflicts between similarly, and highly, attractive concurrent goals and, hence, in many cases to reduce the effects of interference. Critically for the relationship of memory control to anxiety, we postulate that the septo-hippocampal system achieves this resolution by increasing the valence of affectively negative associations of those goals. So, in our view, 'amnesia' resulting from hippocampal damage could more properly be described as 'catastrophic hypermnesia', exemplified by reports of apparent failure to retrieve correct items because of intrusion errors (i.e. retrieval of items that were previously correct but are now incorrect). It is the variety of ways in which the specific methodological features of a memory test can generate conflict between correct and incorrect items that has given rise to the profusion of dichotomies of 'types of memory' supposedly supported or not by the hippocampal formation.

We see the same basic mechanisms, when hyperactive or hypersensitive, as giving rise to generalized anxiety disorder. Increases in the valence of affectively negative associations cause not only current goals to be judged as more threatening than normal, but also a consequential bias towards the storage of more threatening associations of those goals. Thus anxiety has not only an emotional component, but also a unique cognitive component. We shall argue that, while the amygdala is indeed necessary for the expression of many emotional aspects of anxiety, it is so because of its interactions with the septo-hippocampal system; the latter remains the central structure for the specific cognitive components of anxiety and also the principal mediator of the behavioural effects of the anti-anxiety drugs.

1.1 STARTING POINTS FOR THE NEUROPSYCHOLOGY OF ANXIETY

Part of each of our heresies is simply semantic. Our disagreement with some of the orthodox views turns not on the experimental facts (most of which are not in dispute),

but on decisions as to which facts are to be included as explicanda, what inferences are to be drawn from them, and which words should be used to summarize them.

This part of the disagreement arises from a fundamental obstacle in the way of the construction of any neuropsychological account of 'anxiety': the most immediate referent for the state of anxiety is our own experience. As a result, the most direct objective measure of such a state would seem to be the say-so of a human self-observer (who, for example, answers a question in an interview or places a mark on a scale or questionnaire); but the only practicable way to infer causes in the neural control of anxious behaviour (even given the recent advent of neuroimaging techniques for studying the living human brain) is with experimental animals. Somehow, therefore, we have to bridge the gap between the introspections of human beings and the experimental study of animals. That is where the inferences and the semantics come in to cloud discourse.

At least three different starting points have been used by scientists attempting to bridge the gap between human reports of anxiety and animal brain systems. Not surprisingly, these different starting points have given rise to different conclusions.

The most direct starting point depends upon the use of electrical stimulation of the brain to elicit various forms of defensive behaviour. A long line of research (Flynn 1967, 1969; Panksepp, 1982; Graeff, 1991) has demonstrated that it is possible to elicit escape behaviour ('flight') or defensive aggression ('fight') by stimulating points in the medial hypothalamus and the central periaqueductal grey. In the great majority of instances, whether fight or flight is elicited at these points is determined not by the parameters of stimulation, but by environmental conditions. Roughly speaking, if escape is possible, that is the dominant response; if it is not, but something that can be attacked is present in the test chamber, then defensive aggression is seen. As a result, these experiments are often taken as delineating a unified 'fight–flight system'. We do not dispute these observations. But we do not believe that one can take such fight–flight behaviour to be a consequence of a central state equivalent to that of human anxiety. This is especially so since such fight–flight behaviour needs to be categorically distinguished from other classes of defensive behaviour even in rats (Blanchard and Blanchard 1990a,b), and appears sensitive to anti-panic, or 'panicolytic', but not anxiolytic drugs (Blanchard *et al.* 1997).

The second starting point, which has given rise to the current amygdalocentric orthodoxy, is that of conditioned fear. As is well known, if an initially neutral stimulus (e.g. a light flash or a brief tone) is paired, in a standard Pavlovian conditioning paradigm, as a conditioned stimulus (CS) with a painful unconditioned stimulus (UCS) such as electric foot-shock, the CS comes to elicit a variety of conditioned responses (CRs) that can plausibly be interpreted as signs of fear (e.g. freezing/immobility, increased heart rate, defecation, increased readiness to startle, etc.). Extensive experimental work, especially in the laboratories of Joe LeDoux (1992, 1994) and Michael Davis (1992*a*,*b*,*c*), has established that a key site for the formation of such CRs lies in the amygdala. Again, we do not dispute the observations. Our heterodoxy lies, rather, in the belief that conditioned fear of this kind is not equivalent to human anxiety as this presents in the clinic but, rather, is more closely related to phobia.

The third starting point, and the one we favour, lies in the behavioural effects of the anxiolytics, i.e. drugs which reduce self-reported or physician-assessed anxiety in the

clinic. As indicated above, this approach has been rendered much more powerful by the availability of the novel anxiolytics. These share the clinical anxiolytic effect of classical anxiolytics but have different mechanisms of direct pharmacological action and non-overlapping side-effects. Neural or behavioural effects shared by both classical and novel anxiolytic drugs have, therefore, a high probability of delineating processes involved in clinical anxiety or in its amelioration. Our approach is to use the common actions of all clinically well-established anxiolytic drugs as markers for anxiety itself. Remarkably, this emphasis on common action has led to no major changes in the 1982 theory despite its having been based entirely on data from the classical anxiolytic drugs.

Early study of the effects of classical anxiolytics on the behaviour of experimental animals, especially in paradigms derived from animal learning theory, gave rise to the concept of a 'behavioural inhibition system' (Gray 1976; see Fig. 1.1 and, for progressively more complex versions, Figs 3.1 and 5.1). This system is held to control the inhibition of ongoing behaviour, the increase in vigilance, and the increase in arousal which can be produced by stimuli associated with pain, punishment, failure, loss of reward, novelty, or uncertainty. Ever more extensive analyses of relevant lesion data have repeatedly suggested (Gray 1970a, 1977; Gray and McNaughton 1983), and by now have to all intents and purposes established (Appendix 8), that the septo-hippocampal system is a key element in the behavioural inhibition system. This hypothesis was supported by the finding that classical anxiolytics impair the control of hippocampal theta activity. There is now particularly strong evidence in its favour in that the novel anxiolytics, which nowhere figured in the construction of the original theory, have been

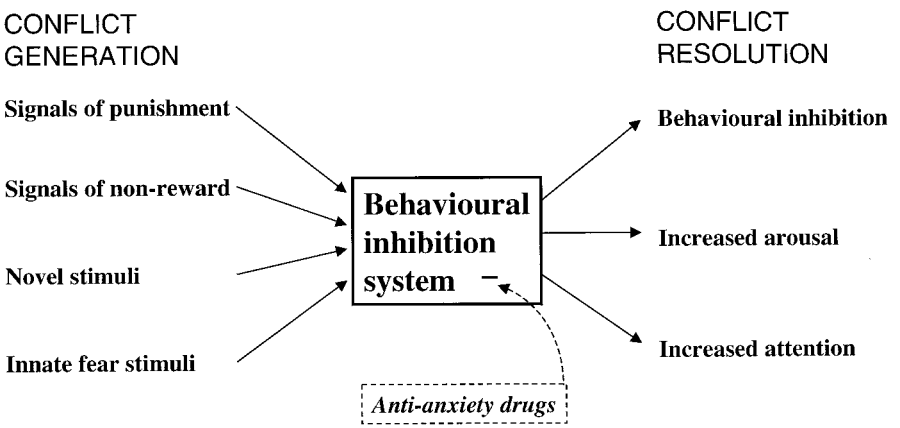


Fig. 1.1 The behavioural inhibition system (Gray 1982). This responds to any of its adequate inputs (CONFLICT GENERATION) with all of its outputs (CONFLICT RESOLUTION). It comprises the hypothetical substrate on which the anti-anxiety drugs act to reduce anxiety (Chapters 2–4; Appendix 1). Note that the key feature of all stimuli which activate the behavioural inhibition system is that they should generate conflict between competing goals. Thus, where a to-be-punished response is weak or where a novel stimulus elicits only approach tendencies uncontaminated with avoidance, the behavioural inhibition system will *not* be engaged.

found also to fit its predictions in having essentially the same effects as classical anxiolytics not only on behaviour[1] but also on the control of hippocampal theta.

A fourth possible starting point (one which, however, appears to have given rise to no major independent theory) lies in the effects of brain lesions in man. In what could be argued are the most extreme cases of anxiety, resistant to both psychological and pharmacological treatment, 'psychosurgery'—lesions of the cingulate or prefrontal cortex—has been used as treatment with some degree of success. So these cortical areas could well mediate extreme (Marks *et al.* 1966) or complex forms of anxiety, especially (as we shall see in Chapter 11) in the case of obsessive–compulsive disorder (Rapoport 1989).

To integrate the different conclusions reached by these different starting points, we have had to resolve one of the thorniest semantic problems in this area: the varying uses of the terms 'fear' and 'anxiety'. Sometimes these are used as synonyms; sometimes, almost as antonyms. While this problem is particularly severe for clinical classifications (Chapter 11), it is one which, more surprisingly, also bedevils behavioural analysis in animals (Chapters 2 and 3). Its resolution is crucial for clear understanding of the relations between the different points of view just described.

1.2 FEAR VERSUS ANXIETY

Since 1982, 'ethopharmacological' analysis (in which ethology rather than learning theory takes centre stage; Blanchard and Blanchard 1990a; Blanchard *et al.* 1993) has produced results which are critical not only for distinguishing 'fear' from 'anxiety' in general, but also for an important development in the current version of our theory. The behavioural inhibition system of the first edition remains central to the model presented here, but ethopharmacological analysis has provided a superior way of defining its inputs. These (see Fig. 1.1) were originally defined quite independently of each other (and in that sense arbitrarily) by the results of pharmacological and behavioural experiments. Ethopharmacological analysis provides a functional basis for this definition which encompasses (with the help of some learning-theory generalization) all of the inputs to the system in a single rubric.

The crucial point (see Chapter 2 for a full discussion) is that the forms of behaviour that are appropriate when, for example, a rat must *leave* an area where there is a cat are quite different from those that are appropriate when a rat must *enter* an area where a cat has been or might be. Cats and cat odour, respectively, have the advantage of being stimuli which can release these two different classes of behaviour without the need for prior training; but it is the form of behaviour that is important, not the type of stimuli which elicit it. Also important is Blanchard's concept of defensive distance.

1. The effects of the novel anxiolytics on both animal behaviour and in the clinic appear to show ∪-shaped dose–response curves; to require long-term administration for clear effects; and to show some other minor differences from classical anxiolytics. We resolve this apparent difficulty (Appendix 1) by appeal to evidence that this is due to their opposite effects on the pituitary–adrenal axis to those of classical anxiolytics. But note that, in any case, the similarity of the animal behaviour and human clinical response in this regard supports rather than weakens the theory.

Different forms of behaviour are appropriate at different distances from the cat (e.g. fight is only appropriate when the cat is close), but this 'intensity' variable is independent of the cat/cat odour distinction. The Blanchards included the cat-related behaviours in a 'fear defence battery' and the cat-odour-related behaviours in an 'anxiety defence battery'. Having already distinguished 'fear' and 'anxiety' on purely theoretical grounds, they then showed that the anxiety but not the fear behaviours are affected by anxiolytic drugs. Consistent with this analysis, simple phobia in people is also insensitive to anxiolytics (Sartory *et al*. 1990).

The Blanchards themselves base this categorical distinction between fear and anxiety on whether the predator is actually or only potentially present. However, learning-theory analysis (Chapter 3) suggests, rather, that the critical issue is whether the behaviour functions to remove the animal *from,* or to facilitate entry *into,* a dangerous situation—i.e. whether active or passive avoidance is involved: in the former case, fear is involved, in the latter, anxiety. This categorical distinction of 'defensive direction' is separate from the smoothly graded dimension of 'defensive distance' mentioned above. Crucially (Chapter 3), it allows us to include in our analysis of fear versus anxiety the events of punishment and omission of reward, as well as stimuli related to these events. The two distinctions, of defensive direction and defensive distance, are theoretically independent of one another: an animal may be engaged in approach to, or escape from, danger at any distance from the point of danger. The closer the animal is to that point, the more intense will be the degree of activation of whichever is the corresponding emotion system and emotional state (anxiety during approach, fear during escape, in the Blanchards' terminology, which we shall, however, increasingly elaborate as the book proceeds). The two dimensions are not, however, in practice independent, since (given a fixed appetitive motivation conflicting with the danger) the closer the animal is to the point of danger, the more likely it is that escape will take precedence over anxious approach. In the straight alley, as exemplified in Neal Miller's classic studies of approach–avoidance conflict (see Gray 1987b, Chapter 9), the animal approaches the goal and then oscillates back and forth at a point where the decrease in defensive distance is such that the approach and avoidance tendencies are just in balance. It should be noted, however, that defensive distance, as defined by the Blanchards, is a cognitive construct. It does depend on physical (or temporal) distance from a danger, but also on the perceived amount of threat: the greater the perceived threat, the greater is the physical distance from it that combines with perceived threat to produce a particular value of defensive distance.

1.3 THE HIERARCHICAL DEFENCE SYSTEM

We are now in a position to address at least part of the problem that, from our four basic starting points, we have arrived at multiple candidates for the 'seat of anxiety': the central grey and hypothalamus; the amygdala; the septo-hippocampal system; and the frontal and cingulate cortices. Can all these inferences be correct? Clearly, each of these regions plays some kind of a role in defensive behaviour and the emotional states that accompany such behaviour. And, since the brain is a single coordinated (albeit highly differentiated) system, it is also reasonable to assume that the regions indicated

above must interact in some systematic fashion in the control of defence in general and, possibly, anxiety in particular. Indeed, much of this book will be devoted to an analysis of just these interactions. But we nonetheless need some way of deciding which, if any, of electrically elicited fight–flight behaviour, conditioned fear, the behavioural effects of anxiolytic drugs or human psychosurgery gives the best clue to the neural basis of anxiety.

Ethological analysis has provided us with a view of a range of behaviours, each elicited by some specific combination of defensive distance and defensive direction. Both behaviour and stimulus analysis can thus range from 'slow and sophisticated', where there is room and time to manoeuvre, to 'quick and dirty' (LeDoux 1994), when danger is immediate. These functional requirements, and more detailed neural analysis (Chapter 6), have led to the view of a hierarchical defence system (e.g. Graeff 1994) in which (Fig. 1.2) the lowest level, the periaqueductal grey, coordinates undirected escape; the medial hypothalamus coordinates directed escape; the amygdala coordinates simple active avoidance; and the anterior cingulate coordinates more complex active avoidance.

Matching this hierarchical animal-based view is clinical evidence which now demonstrates—as it did not when the first, 1982, edition of this book was prepared—that, underlying the whole gamut of neurotic disorders, we need to postulate at least two separate neuropsychological systems (see Chapter 11). The critical evidence takes the form of a pharmacological double dissociation: there are drugs which alleviate self-reported anticipatory anxiety but not panic attacks, and others which do the reverse—they alleviate panic but not anxiety (Klein 1980; see also Blanchard *et al.* 1997). This means that there are now two jobs for which our candidate brain regions can apply: as the 'seat of anxiety' or as the quite separate 'seat of panic'. The choice for the second of these two jobs has been made relatively easy by the careful analysis which has been

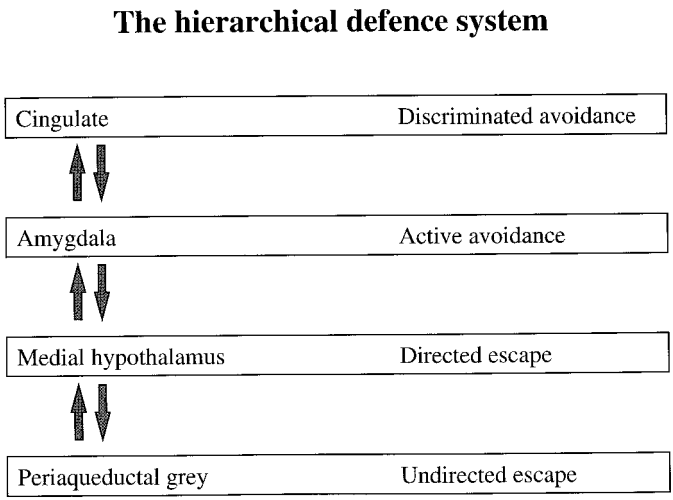


Fig. 1.2 A simplified view of the structures controlling active defence illustrating the generally hierarchical organization of defence systems. Progressively more central structures in the brain (periaqueductal grey–cingulate) are held to deal with progressively more complex aspects of defence (undirected escape–discriminated avoidance).

undertaken recently of the points in the mammalian brain from which fight–flight behaviour can be elicited (see Graeff 1994). The effects of such stimulation map well onto autonomic changes observed in human beings undergoing a panic attack, and the fight–flight behaviour itself is sensitive to panicolytic but not anxiolytic drugs (Blanchard *et al.* 1997). Thus the hypothalamic–central grey axis is the preferred joint candidate for the brain system underlying panic. Furthermore, anxiolytics not only do not have effects comparable to those of panicolytic drugs, but actually increase the degree of fight–flight behaviour elicited by stimulation of the central grey (Graeff 1994). These observations support the inference from the clinical data that anxiety is not the same state as panic, and indeed go beyond it to suggest that anxiety may actually inhibit panic. (Our full theory, however, includes the postulate that panic can also be a symptom of extreme levels of anxiety which override this concurrent inhibition; see Chapter 11.)

We can, therefore, allocate the central grey and related hypothalamic regions to panic. The role of the amygdala in fear, as shown in experiments exposing animals to pain-associated CSs, has been confirmed in human studies of patients with amygdalar lesions (Calder *et al.* 1996) or using neuroimaging methods (J. Morris *et al.* 1996; Phillips *et al.* 1997, 1998), which show amygdalar involvement in the response to expressions of fear in both facial and auditory modes. This, then, is the obvious choice of key structure in relation to the specific phobias (which need to be carefully distinguished from more complex states involving anxiety, as discussed in Chapter 11). This leaves us with the major question to which this book is addressed: what neural structures mediate the neuropsychology of anxiety proper? Since the dominant view at present favours the amygdala for this choice, we start with the reasons why we believe this choice to be largely wrong. (However, in our final theory, the amygdala continues nonetheless to play a substantial role in the neuropsychology of anxiety.)

1.4 THE AMYGDALA

Our first objection is based on the fact that, in a sense, the amygdala does too much (Chapter 6). For a start, it appears to be responsible for the formation of cue–reinforcer associations not only when the reinforcer is aversive, but also when it is appetitive. Thus, lesions to the amygdala impair not only conditioned fear, as measured, for example, by cardioacceleration, but also active avoidance and rewarded approach behaviour. Unconditioned responses such as aggressive and sexual behaviour also depend to a major extent on the amygdala. Given these data, we accept that the amygdala plays a major role in controlling the output of many, perhaps all, emotion systems. Indeed, as we shall see, we view output to the amygdala from the septo-hippocampal system as crucial in generating the arousal and autonomic components of anxiety. But there is little reason to attribute to the amygdala a role *specifically* in anxiety. The fact that it is involved in controlling autonomic output and behaviour for so many different classes of emotional stimuli argues that it is, rather, a general controller of emotional output of all types.

Second, a key feature of human anxiety lies in apprehension of the possibility of, not pain, but failure or loss of reward. There are a number of well-controlled experimental paradigms in which such 'anticipatory frustration' (Amsel 1962, 1992) can be measured

in animals; and there is much experimental evidence that anxiolytics weaken this state (Gray 1977). Studies of behaviour in these paradigms after lesions to the amygdala, however, have usually failed to show the effects that would be expected if such lesions reduce anxiety (Appendix 2).

Lastly, if one does accept the anxiolytic drug starting point for the analysis of the neuropsychology of anxiety, there is a critical lack of concordance between the behavioural effects of these drugs and lesions to the amygdala, respectively. Amygdalar lesions have failed to reproduce the effects of anxiolytic drugs on behaviour in a number of schedules involving reward omission, on two-way active avoidance, and on rearing (this last being one of the key components of the Blanchards' anxiety defence test battery). If the amygdala is the 'seat of anxiety', some explanation for this lack of concordance is required.

Despite this negative conclusion, we accept that a full account of the neuropsychology of anxiety requires adequate consideration of the roles played by the amygdala and especially of the interactions between this structure and our preferred candidate, the septo-hippocampal system, to which we now turn.

1.5 THE SEPTO-HIPPOCAMPAL SYSTEM

The septum and the hippocampus are strongly interlinked and are separated from much of the rest of the brain by the ventricles. As a result they can fairly easily be removed from the brain by blunt dissection (Fig. 1.3A). This feature of the gross anatomy gives us the first of many reasons for speaking of a unified 'septo-hippocampal system' (Fig. 1.3B). The evidence for the involvement of this system in the action of the anxiolytic drugs is virtually the mirror image of the case against the amygdala. In this brief overview, we emphasize three points.

First, the parallel between the behavioural effects of the anxiolytic drugs and those of experimental lesions to the septo-hippocampal system is strikingly strong (see Chapter 4 and particularly Table 4.1). We speak here of lesions to the 'septo-hippocampal system' because the parallel in fact has three limbs. We first pick out those effects of lesions to the hippocampal formation that are matched by similar effects of lesions to those regions of the septal area with which the hippocampus is most closely connected, namely, the medial septal area, which projects to the hippocampus, and the lateral septal nuclei, which receive a return projection from the hippocampus. These joint effects of hippocampal and septal lesions then turn out to resemble closely those of anxiolytic drug administration.

This parallel was first pointed out by Gray (1970a) on the basis of a then very limited data set confined largely to the barbiturate sodium amytal (also known as sodium amylobarbitone and sodium amobarbital) and a small number of lesion experiments. It then proved robust when tested against a much wider data base taken from experiments (Gray 1977) on all the classical anxiolytic drugs (barbiturates, benzodiazepines and ethanol) as well as several hundred lesion studies (Gray and McNaughton 1983). The emergence since 1983 of data on the novel anxiolytics, especially antagonists at the 5-hydroxytryptamine (5HT) 1A receptor such as buspirone, has provided a further critical

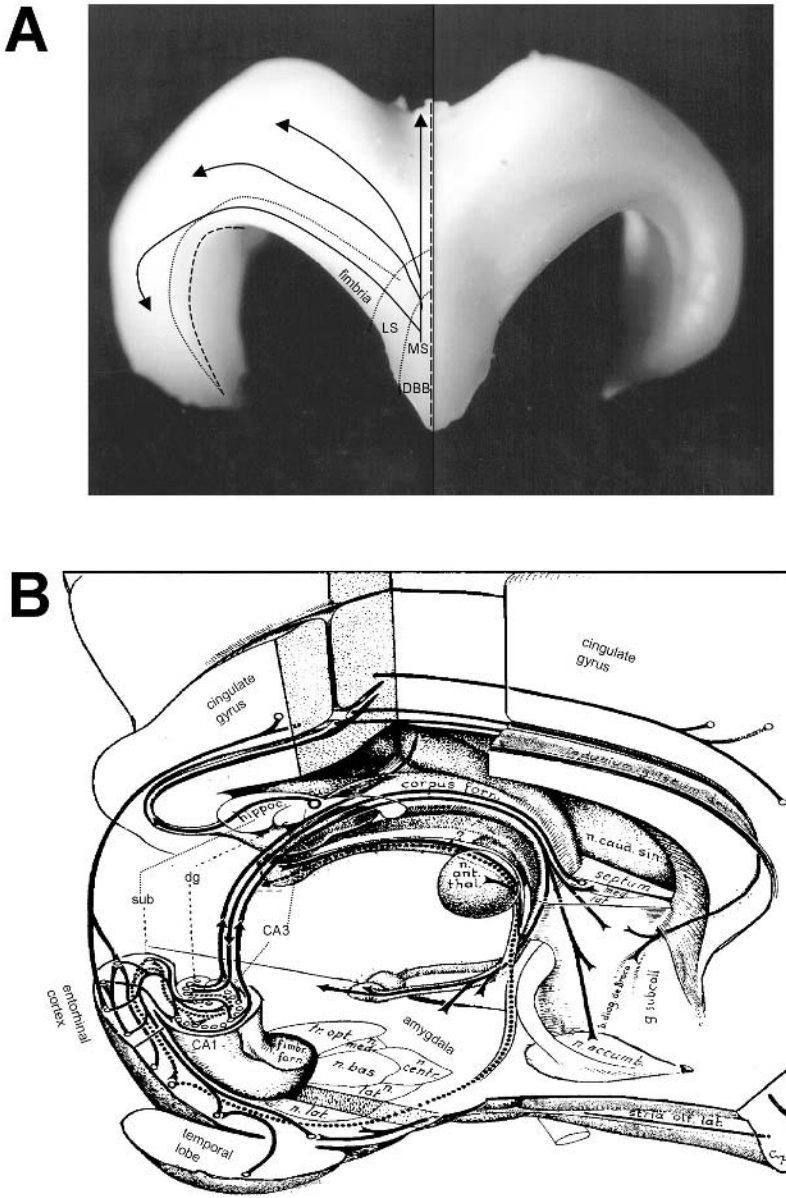


Fig. 1.3 The hippocampal formation and its relation to the septal area. (A) The hippocampal formation and septal area dissected from the brain of a rat. (Photograph by J. P. Broad, from Rawlins 1977.) We have roughly indicated the location of the medial septum (MS), diagonal band of Broca (DBB), the lateral septum (LS), and the fimbria. (B) The hippocampal formation and septal area drawn as if partially dissected out of the brain of a primate. (From Gastaut and Lammers 1961.) The arrangement is essentially as in A except for the elongation of the fornix which now clearly separates the hippocampus from the septal nuclei. A segment has been cut out of the middle of the hippocampus (hippoc.) showing the separate cell fields of Ammon's Horn (CA1, CA3), the subiculum (sub), and the dentate gyrus (dg).

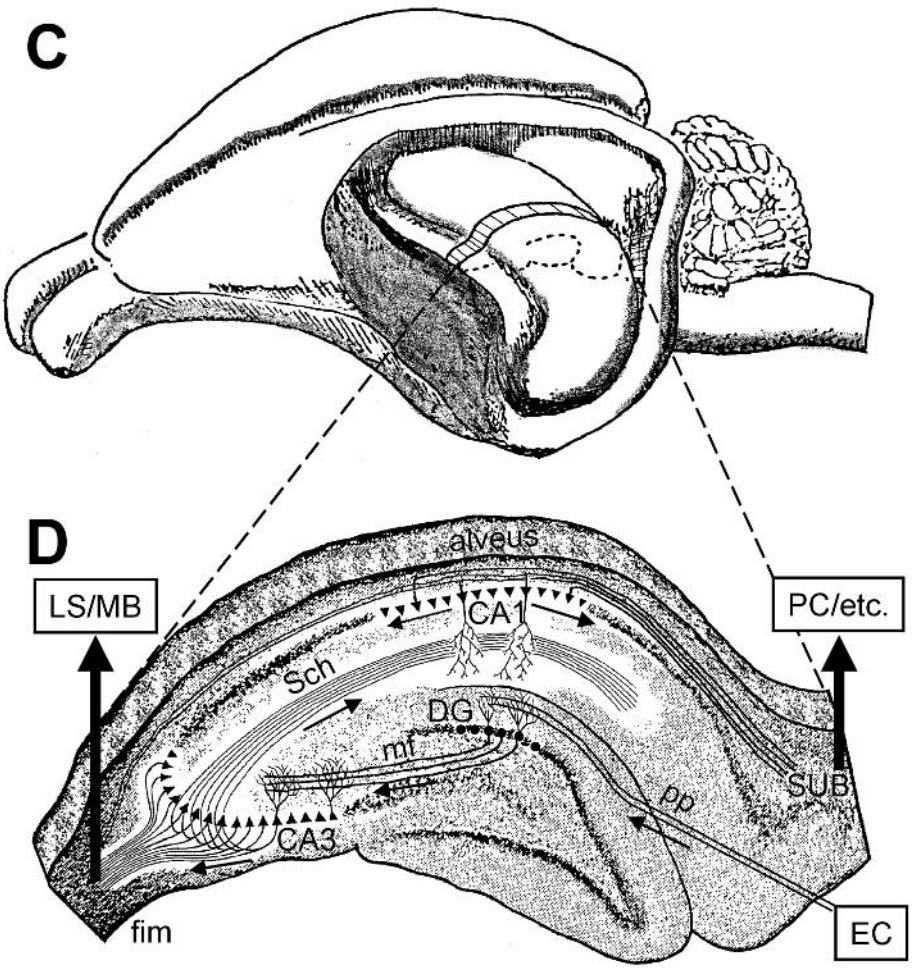


Fig. 1.3 (C) A roughly equivalent dissection of the rabbit brain with the parietal and temporal neocortex removed to expose the hippocampal formation. This shows a nominal lamellar slice through the middle of the hippocampus of the type which can be removed for *in vitro* physiology and which contains essentially the same circuitry and has cells in essentially the same configuration in humans. (Modified by N. McNaughton from Andersen *et al.* 1971.) (D) Primary circuitry of the mammalian hippocampus. This is essentially repeated in each slice (lamella). Input from the medial and lateral entorhinal cortex (EC) enters in the medial and lateral perforant paths (pp) which synapse in middle and outer portions, respectively, of the dendrites of the dentate gyrus (DG) granule cells. These send mossy fibres (mf) which synapse on CA3 pyramidal cells which, in turn, send output to the lateral septum (LS) and, via the Schaffer collaterals (Sch), to the CA1 pyramids. Area CA1 sends a weak projection, via the alveus and fimbria (fim), and a stronger projection to the subiculum (SUB) which, in turn, sends projections to areas such as posterior cingulate cortex (PC) and, via the alveus and fimbria, to the mammillary bodies (MB). (Modified by N. McNaughton from an original drawing by D. S. Kerr.)

test. These compounds have effects on animal behaviour which fit into the same model of the behavioural inhibition system (Fig. 1.1) that was able to encompass the effects of the classical anxiolytics (Appendix 1). So, this cornerstone of our theory not only has an immense volume of data supporting it but has shown predictive validity for more than a quarter of a century, and across drug types with no known common direct actions on the nervous system and no common clinical effects other than the relief of anxiety.

Second, experiments from our laboratories using the classical anxiolytics had demonstrated prior to the first edition that, despite their diverse chemical structures, these all have in common two electrophysiological 'signatures' that involve hippocampal theta activity—signatures that, taken together, are not shared with any drugs shown not to be anxiolytic in the clinic.

As noted briefly above, theta activity is a rhythmic burst firing pattern of neurons in the hippocampus and related structures which, because it is synchronous across very large numbers of cells, often gives rise to a high-voltage quasi-sinusoidal electrographic slow 'theta rhythm' (approximately 5–10 Hz in the unanaesthetized rat) that can be recorded from the hippocampal formation under a variety of behavioural conditions. Theta activity is controlled by pacemaker cells located in the medial septal area, which themselves receive a frequency-setting input from structures in the midbrain, including importantly the medial supramammillary nucleus (Kirk and McNaughton 1991, 1993). It is possible to drive the theta rhythm by low-frequency electrical stimulation of the medial septal area (one theta wave being elicited by each pulse) or to elicit it by high-frequency (e.g. 100 Hz) stimulation of the reticular formation (which provides input to the medial supramammillary nucleus, where frequency is encoded and relayed to the medial septum).

The common electrophysiological signatures of the classical anxiolytics, demonstrated prior to the first edition of this book, consist in: (1) an elevation of the threshold for septal driving of hippocampal theta that is relatively specific to an intermediate theta frequency of 7.7 Hz (at which, in the undrugged male rat, there is a minimum threshold for theta driving; Fig. 9.1); and (2) a reduction in the frequency of theta elicited by reticular stimulation of given intensity (in both drugged and undrugged animals there is a linear relation between stimulus intensity and elicited theta frequency with no non-linearity in the region of 7.7 Hz; Fig. 9.2). These signatures are found for all the classical anxiolytics, and we have been unable to demonstrate both of them together in any other class of drugs (Chapter 9). As with the drug–lesion parallel noted above, the emergence of the novel anxiolytics offered a stringent test of the predictive generality of these electrophysiological signatures. This test, too, has been passed (Chapter 9; Appendix 5), even in the case of imipramine, an antidepressant which has been shown to have additional action on anxiety independent of its antidepressant effects.

Third, selectively reproducing parts of the septo-hippocampal electrographic signatures of the anxiolytic drugs reproduces parts of the behavioural profile which they share with septal and hippocampal lesions. This is an important result: it demonstrates that the electrographic signatures are not just an accidental correlate of clinical action, but represent functional effects of the drugs on the septo-hippocampal system which can, in combination, reproduce large parts of the behavioural profile of systemic administration of anxiolytics.

In experiments described in the first edition of this book we were able to show that the first, septally elicited, electrophysiological signature of anxiolytic drug action noted above could be reproduced by selective lesion of the dorsal ascending noradrenergic bundle (Gray *et al.* 1975). This pathway arises from the locus coeruleus in the brain stem to innervate most of the forebrain, including the septo-hippocampal system. On the basis of these observations we proposed that the action of the anxiolytics on septo-hippocampal function is in part mediated by drug-induced reduction of activity in the dorsal bundle, an effect that had been directly demonstrated earlier by Fuxe *et al.* (1975). For somewhat similar reasons we additionally suggested that anxiolytic drug action is mediated in part also by reduction in activity in the ascending serotonergic pathway that arises in the median raphe nucleus and innervates the septo-hippocampal system. For both of these monoamine pathways, Fuxe *et al.* (1975) had demonstrated that stress increases release of the relevant transmitter (noradrenaline or serotonin) in the terminal areas, and that anxiolytic drug action, while not affecting basal release, antagonizes this stress-induced increment in release. Furthermore, other experiments, especially from Segal's (Segal and Bloom 1974, 1976) laboratory, had shown that such increases in monoamine release facilitate the passage of neural information around hippocampal circuits. Thus, by counteracting these monoaminergic effects, the anxiolytics would be expected to impede hippocampal information processing.

The behavioural effects of lesions of the dorsal ascending noradrenergic bundle were interesting (McNaughton and Mason 1980). For many tasks they did in fact reproduce the effects of anxiolytic drugs, essentially in full. But, for a substantial number of tasks sensitive to anxiolytics and septal and hippocampal lesions, they were totally without effect. The effects of lesions or other disruption to the ascending serotonergic pathway appeared to fill in some, but not all, of the gaps, giving rise in particular to behavioural disinhibition in tasks involving exposure to painful stimuli (whereas lesions of the dorsal noradrenergic bundle have greater effects in tasks involving stimuli associated with non-reward).

In a further narrowing of the gaps, we have recently mapped parts of the ascending system controlling the frequency of theta, the source of our second, reticular-elicited, electrophysiological signature of anxiolytic action. We found that this could be reproduced by injections of a benzodiazepine into the medial supramammillary nucleus (McNaughton *et al.* 1995). This is a small nucleus which relays information arising in areas such as the nucleus reticularis pontis oralis via the medial forebrain bundle to the medial septum; and is, we have shown, one site at which theta frequency is actually computed (Kirk and McNaughton 1991, 1993). We have now demonstrated that, in a fixed interval task, involving non-reward but insensitive to dorsal noradrenergic bundle lesions, medial supramammillary injections of a classical anxiolytic both reduced theta frequency and affected behaviour to the same extent and in the same way as did systemic injections (Senior *et al.*, in preparation).

Putting together these new observations with the earlier ones, our theory now holds that anxiolytic drugs directly reduce the release of noradrenaline and serotonin into the septo-hippocampal system and directly alter the encoding of theta frequency via areas such as the supramammillary nucleus. These direct effects on modulatory systems then degrade (but do not totally eliminate) information processing in the septo-hippocampal

system. In turn, the changes in septo-hippocampal function then account for the bulk of the behavioural effects of all clinically effective, centrally acting, anxiolytic drugs. However, some further effects of many of these drugs (especially, as we shall see, on the fear-potentiated startle response) are achieved by an additional direct action on the amgydala.

1.6 WHAT IS THE SEPTO-HIPPOCAMPAL SYSTEM?

What we have so far rather loosely presented as a hypothesis of anxiolytic drug action that focuses on the septo-hippocampal system is therefore more properly described as a hypothesis that postulates actions of these drugs on ascending noradrenergic, serotonergic, and GABAergic pathways. These actions block increases which would otherwise have occurred in the signal-to-noise ratio of hippocampal processing. (The amines affect the signal-to-noise ratio of spatial summation in neurons, while the theta-frequency-controlling GABAergic pathway affects the signal-to-noise ratio of temporal summation; see Appendix 5.) These various subtle changes in the control of hippocampal theta activity (and hence often in the gross theta rhythm) result in a reduced acuity of the septo-hippocampal system in its processing of threat signals (and of negative affective valence in memory paradigms).

This central role of theta activity in our theory of anxiolytic drug action gives rise to our definition of the septo-hippocampal system itself, and explains the fact that it does not completely overlap with what is anatomically the hippocampus proper. For we treat as 'septo-hippocampal system' all those structures which receive direct, theta-activity-controlling, inhibitory GABAergic input from the medial septal area. The septo-hippocampal system, on current knowledge, therefore includes the hippocampus proper, the dentate gyrus, the entorhinal cortex, the subicular area, and the posterior cingulate cortex. The roles in behaviour of all these regions, as well as the ascending systems that control the theta rhythm, are therefore relevant and will accordingly be reviewed in the chapters that follow. In these reviews we shall attempt to strengthen and refine the central parallel, already noted, between, on the one hand, the behavioural effects of lesions to the septal area and hippocampus and, on the other, those of the anxiolytic drugs.

While the septo-hippocampal system, so defined, is anatomically much more extensive than the hippocampus proper, there are two important ways in which the effects of anxiolytic drugs are likely to be more limited than the list of structures comprising the septo-hippocampal system might imply. First, the anxiolytic drugs do not impair all aspects of the control of theta (for example, they do not produce the state-dependent total blockade of theta which is characteristic of reduction in the cholinergic input; Appendix 5). They are likely, therefore, to produce only a subset of the behavioural effects which can be produced by total elimination of theta activity. For example, as noted above, the effects of destruction of the dorsal ascending noradrenergic bundle appear to cover only a subset of the effects of lesions to the septo-hippocampal system. Second, the anxiolytic drugs, even if they eliminated theta activity, would not produce a complete dysfunction of the septo-hippocampal system. Because, according to our model (in this second edition), they only affect theta activity (and not other aspects of

septo-hippocampal processing), they only degrade rather than eliminate processing in the septo-hippocampal system. Hence, when they have effects like those of septal or hippocampal lesions, we would expect these effects to be weaker. So, taking these two points together, we would not expect the central parallel between lesions to the septum and hippocampus proper, on the one hand, and anxiolytic drug treatment, on the other, to be qualitatively or quantitatively exact. What is surprising, under these circumstances, is how very good the parallel in fact is.

Nonetheless, although we suppose the action of anxiolytics on the septo-hippocampal system to be indirect and incomplete, it still follows from the centrality of this action that an account of the neuropsychology of anxiety must be derived from an account of the cognitive functions of the septo-hippocampal system, and particularly the hippocampal formation. We turn now therefore to consider this, our other heresy: the notion that the functions of the hippocampal formation relate to anxiety (as distinct from spatial cognition or memory).

1.7 WHAT DOES THE SEPTO-HIPPOCAMPAL SYSTEM DO?

Since the pioneering insight of O'Keefe (O'Keefe and Dostrovsky 1971), based on observations of just seven single neurons in the hippocampus of the freely moving rat, it has been confirmed repeatedly that, when the rat is allowed to traverse a spatially extended environment, such neurons display 'place fields', that is, they tend each to fire systematically in one part of the environment but not others. These observations led O'Keefe and Nadel (1978) to propose that the key cognitive function of the hippocampus is to process spatial information. Their hypothesis proposed also that the hippocampus forms and stores spatial memories, as recently reaffirmed by Nadel (1991). Recent data suggest that the hippocampus does not provide a long-term store of any type of information and so we focus here on the spatial cognition, rather than memorial, aspects of the O'Keefe and Nadel (1978) model.

In essence, the function these workers attributed to the hippocampus is that of forming, in subhuman animals, spatial maps. At the human level, it was proposed that this spatial mapping faculty becomes a more general faculty of cognitive mapping, applicable to sensory and conceptual domains of all kinds. We focus here on the conceptually simpler spatial aspects of the O'Keefe and Nadel model, confining the discussion accordingly to subhuman animals. The evidence from lesion experiments has provided overwhelming support for the following proposition: tasks that require animals to solve problems by navigating in spatially extended environments, e.g. mazes or Morris's (1981, 1984) widely used swimming pool, are especially sensitive to the deleterious effects of lesions to the hippocampal formation, or to only the hippocampus proper, and usually more so than apparently equivalent non-spatial tasks (Morris *et al.* 1982, 1990; Jarrard 1983). Thus, that the hippocampus has a key role of some kind to play in the solution of spatial tasks, consistent with the O'Keefe and Nadel hypothesis, is abundantly clear. It is also relevant that both novel and classical anxiolytics have qualitatively similar effects to hippocampal lesions in this regard (McNaughton and Morris 1987, 1992), and that even mild selective disturbance of the frequency of theta activity by the injection of a

classical anxiolytic into the medial supramammillary nucleus can produce spatial deficits (Pan and McNaughton 1997). Given the central role of theta activity in O'Keefe and Nadel's model, these latter results are also consistent with their view.

What is not so clear is whether, as postulated by O'Keefe and Nadel, the functions of the hippocampus are exclusively (or even directly) to do with spatial cognition. The earliest literature contains numerous instances of clearly non-spatial cognitive deficits in animals after hippocampal damage. But it was usually possible to discount these reports on the grounds that the damage extended well beyond the hippocampus proper into other components of the hippocampal formation, or even regions outside the latter altogether. (Note, however, that the septo-hippocampal system as defined by our theory encompasses most of the structures which would have been incidentally damaged by old style 'hippocampal' lesions.) That rebuttal of such reports, however, is no longer possible: there are clear instances of damage limited to the hippocampus proper which has led to deficits in non-spatial performance. Two examples of this kind will suffice. First, Davidson and Jarrard (1993) studied rats with excitotoxic lesions which removed virtually all the hippocampus proper (including the CA1–CA3, dentate granule, and hilar regions) but nothing else. These animals were deficient in bringing instrumental behaviour under the control of a clearly non-spatial contextual cue, namely, the state of hunger. Second, Ridley *et al.* (1995) studied marmosets with excitotoxic lesions confined to the CA1 pyramidal cell field of the hippocampus proper. These animals were impaired in conditional but not simple discriminations to an equal extent, whether the dimension along which the discrimination was required was spatial (position) or not (object irrespective of position). Consistent with these findings from lesion experiments, we have shown recently that virtually all of several hundred cells recorded from the hippocampus proper and entorhinal cortex had task-related correlates in a non-spatial timing task (Young and McNaughton 1997). Most of these 'timing fields' were unique and virtually none was consistent with the coding of a 'place field' by the cell in question.

A second issue is whether the 'spatial' aspects of hippocampal functions are truly spatial. It has recently been shown that rats with hippocampal system damage can learn and remember spatial positions perfectly well provided training or testing is conducted suitably (Eichenbaum *et al.* 1990; Whishaw *et al.* 1995; Whishaw and Tomie 1997; Hanneson and Skelton 1998). This suggests that the problems in spatial tasks could have a non-spatial origin. Similarly, in Appendix 6 (and see Chapter 7), we review extensive data which suggest that 'place fields' do not code unambiguously for places, change with non-spatial changes in the environment, and are not topographically mapped into the hippocampus. All of these data suggest that the 'placeness' of fields is not the result of uniquely spatial encoding. They also demonstrate that no structure receiving output from the hippocampus could, from that output alone, determine unambiguously and at all times where the animal is in space.

The various non-spatial memory hypotheses of hippocampal function grew out of observations completely different from those that inspired the spatial hypothesis: clinical observations on human beings with temporal-lobe damage associated with profound amnesia, starting with the famous patient, H.M. (Milner *et al.* 1968). Although such damage almost invariably includes many structures besides the hippocampus, this early on became the leading candidate for mediation of the memorial functions compromised

in these patients. The resulting memory hypothesis of hippocampal function has since gone through a number of metamorphoses (see Chapter 8). Two of the earlier views—that the hippocampus plays a major role in the consolidation of memory (i.e. in the transition from short- to long-term storage in general) and that long-term memories are actually stored in the hippocampus—are now largely discredited. But both the clinical and the experimental animal data continue to show that the hippocampus does indeed play an important role in at least some memory tasks. Recent data suggest, furthermore, that damage limited to parts of the human hippocampus proper (and especially the CA1 region) may be capable of giving rise to a profound and enduring amnesic syndrome (Zola-Morgan *et al.* 1986). Equally clearly, however, not all memory functions are compromised by hippocampal damage, or even by the kind of extensive temporal-lobe damage sustained by H.M.

Observations of this kind, showing impairment in some memory tests alongside preservation of memory in others, have spawned a bewildering variety of theoretical dichotomies between hippocampal-dependent and hippocampal-independent forms of memory (Eichenbaum *et al.* 1994). Some members of this list of dichotomies appear to be largely the same as others, apart from the names used, but in other cases real theoretical divergences are involved (Chapter 8). Even in the latter cases most are still current. The hippocampus, then, appears to do something related to some specific type of memory. But, as we did with the spatial hypothesis, we can ask of the memory hypothesis: is this all that the hippocampus does? And, is its involvement in memory tasks due to a specifically memorial function? The evidence from lesion studies is clear: deficits are seen after damage to the septo-hippocampal system in a great variety of tasks that lack any significant memorial component (Gray and McNaughton 1983). As with other aspects of the various controversies concerning hippocampal function, it can be objected to the great majority of these studies that the damage included much besides the hippocampus proper. However, there are again reports in which lesions confined to the hippocampus proper have given rise to non-memorial changes in behaviour. In one such experiment (Jarrard *et al.* 1986) rats with excitotoxic lesions of the hippocampus showed increased resistance to extinction of a running response rewarded with food in the straight alley. Since the same animals had shown no difficulty in acquiring the running response, this increased resistance to extinction could not be attributed to a memorial (or, indeed, spatial) deficit.

Conversely, careful analysis of memory-oriented experiments (Chapter 8) suggests that those aspects of a memory task that are regarded by the theoretically minded experimenter as defining its critical class membership (e.g. working vs. reference or procedural vs. episodic memory) are not what give rise to hippocampal sensitivity. Rather, in many cases, sensitivity to hippocampal damage is conferred by the combination of a source of interference in the task *and* a capacity in the control animals to overcome that source of interference (essentially the position taken by Weiskrantz and Warrington 1975). Taking this together with the fact that amnesics' errors are often intrusions of items which would have been correct in previous trials brings us to the view that hippocampal damage does not prevent subjects from storing correct information. Rather, it prevents them from inhibiting storage or retrieval of incorrect information, which then replaces what would have been a correct response on testing. It is for this reason that we have suggested above that 'amnesia' could be more correctly termed 'catastrophic hypermnesia'.

In addition, as already indicated, the lesion literature demonstrates striking parallels to the behavioural effects of the anxiolytic drugs. This parallel, too, on at least some fronts extends to experiments in which the lesion is confined to the hippocampus proper, as in the Jarrard *et al.* (1986) report of increased resistance to extinction (though the literature on selective hippocampal lesions is too sparse to come to any general conclusion on this score). If we ask of this 'anxiolytic-like syndrome', what septo-hippocampal function does it point to, the answer is necessarily (given the chain of argument we are following) the mediation of some process which is fundamental to anxiety. More specifically, given Fig. 1.1 (deduced from the behavioural effects of the anxiolytics), the septo-hippocampal system is seen as responding to stimuli associated with punishment and non-reward (i.e. to threat) by interrupting ongoing behaviour so as to allow information gathering (the 'attention' output of the behavioural inhibition system). The lesion literature gives ample justification for this view of septo-hippocampal function (Gray and McNaughton 1983). In particular, there are many reports after either septal or hippocampal damage of behavioural changes which are best subsumed as showing a reduction in the capacity for behavioural inhibition (the Jarrard *et al.* 1986 study of extinction of rewarded running is again a good example) and which do not yield easily to an account in terms of impairment in either spatial cognition or memory. Thus, whatever function the hippocampus discharges, its disruption affects spatial cognition, memory, *and* behavioural inhibition, the precise outcome apparently depending largely upon the task the animal is set.

We see as a major virtue of our theory that it attempts to encompass all of these data within a single explanatory framework.

1.8 A COMMON COMPUTATIONAL CORE FOR SEPTO-HIPPOCAMPAL FUNCTION—MARK 1

The anatomy of the hippocampus proper is relatively simple (Fig. 1.3C,D), essentially the same basic circuit being repeated over the entire septo-temporal length of this structure in a series of quasi-strips known as lamellae (Appendix 4). It is not plausible, therefore, to suppose that, buried within it, there are (at least) three different subsystems, one subserving spatial cognition, one some specific class of memory, and one behavioural inhibition. Many workers have therefore proposed that some single core computational function is discharged by the hippocampus, this then being put at the disposal of different types of input–output transaction (involving *inter alia* space, memory, behavioural inhibition, etc.) depending upon the task the animal is required to solve (for a recent example see Eichenbaum *et al.* 1994). Our own approach is along the same lines. The main distinction is that, whereas most other theorists see the hippocampus as carrying out some desirable computation within its own structure, we see it as primarily suppressing undesirable computations in other structures.

There have been significant adjustments to the detailed mechanism of the theory of the first edition to bring it into line with more recent data. Given the incompleteness of our information (Appendix 5) on even the subcortical control of theta (a key aspect of our theory), we expect considerable further revision to be required should there

be a third edition. But the high level descriptors used in the first edition are still useful as heuristic terms which allow easy (if only approximate) appreciation of the scope of our theory. As will be seen, they can largely be viewed as special cases, or aspects, or emergent properties, of the more general and fundamental computational properties ascribed to the septo-hippocampal system in this second edition. So, we will describe septo-hippocampal function first as we did in the first edition, and then from the point of view of the present edition. The main overarching change is that many complex functions we originally ascribed to the septo-hippocampal system itself we now ascribe to its interaction with other structures.

The fundamental suggestion made in the first, 1982, edition of this book was that the septo-hippocampal system acts as a comparator. The key computational functions of the system proposed at that time are sketched out in Fig. 1.4; its presumed anatomical equivalents are outlined in Fig. 1.5. On the basis of (1) the current state of the perceptual world, as assessed by neocortical sensory systems at time t and communicated to the comparator by way of the input to the hippocampal formation from the entorhinal cortex, (2) the animal's ongoing motor programs, as communicated to the comparator by motor programming circuits in the basal ganglia and frontal cortex, and (3) memory stores (comprising stimulus–stimulus, stimulus–response, and response–response associations and located largely in the temporal lobes) of past regularities of experience under similar conditions, the comparator (comprising the hippocampal formation, the subicular area, and the Papez loop that circulates information from the subiculum to the mammillary bodies, the anteroventral thalamus, the cingulate gyrus, and back to the subiculum) computes a prediction as to the next likely state of the perceptual world at time $t + 1$ (the time base between t and $t + 1$ being about 100 ms). Comparison (proposed as taking place in the subiculum) of this prediction with the state of the perceptual world at time $t + 1$ is then the basis for two decisions: match, i.e. the

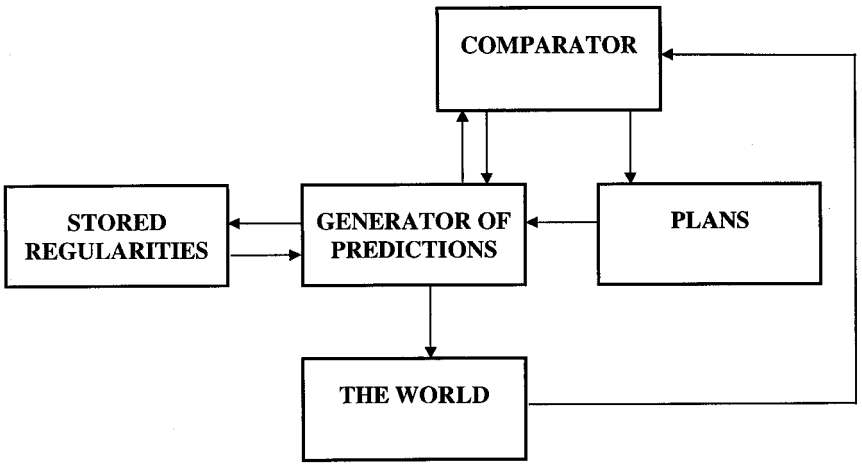


Fig. 1.4 The kinds of information processing required for the successful functioning of the comparator hypothesized in the first edition. (See text for further explanation.)

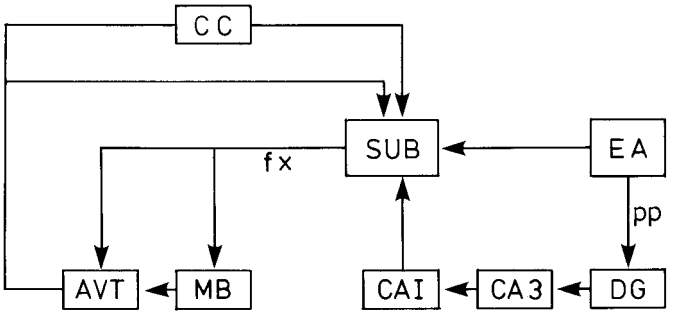


Fig. 1.5 Connections of the subicular area which could allow it to function as the comparator hypothesized in the first edition. AVT, anteroventral thalamus; CC, cingulate cortex; DG, dentate gyrus; MB, mammillary bodies; SUB, subiculum; fx, fornix; pp, perforant path.

prediction is confirmed; or mismatch, i.e. something predicted has failed to occur or something unpredicted has occurred. Since these computations are conducted upon a whole set of selected features of the perceptual world, the final output consists of a vector of such decisions.

The comparator system was seen as operating in two modes. In 'just checking' mode successive outputs are essentially in line with prediction; under these conditions, control of behaviour is left with motor programming circuits and the comparator system carries on monitoring the outputs of the motor program. In 'control' mode, the comparator detects either mismatch or threat (i.e. a predicted input of pain, punishment, or non-reward); under these conditions, the comparator takes control of behaviour and operates the outputs of the behavioural inhibition system as illustrated in Fig. 1.1. Thus the comparator function provided the cognitive or computational heart of the behavioural inhibition system. Jointly, these two concepts were able in the first edition of this book to provide a reasonable account of both the behavioural and the subjective features of clinical anxiety. In addition, they were applied in outline form to the spatial and memorial aspects of hippocampal function.

In essence, this account supposed that, having detected mismatch or threat, the comparator system then attempts, in control mode, to solve the problem thus posed. It does this by gathering further information, both from the environment and from memory stores, in search of alternative descriptions of the perceptual world and/or alternative motor programs that can account for the mismatch or circumvent the threat. In this way, the comparator might identify contextual differences between, for example, the description that gave rise to the prediction and the description of the actual state of the perceptual world in which this prediction has been disconfirmed; or between the motor program that results in exposure to the threat and one that might avoid it. During this period of intensive cognitive processing, the previous ongoing motor program is interrupted (the behavioural inhibition proper output of the behavioural inhibition system) and the animal may be seen to engage in 'risk assessment' behaviour (the 'increased attention' output of the behavioural inhibition system), for example, in the rat, adoption of a 'stretched attend' or 'rearing' posture (Blanchard and Blanchard 1990b); these are precisely the forms of behaviour that are most sensitive to the action of the

anxiolytic drugs. The result of the cognitive processing initiated in this way by mismatch/threat consists in a form of contextual tagging, along the lines: given stimulus X in environment M, action pattern A is correct, but given stimulus Y in environment N, action pattern B is correct.

Application of the notion of contextual tagging to the spatial and memorial aspects of hippocampal function is then relatively straightforward. Note, however, that one of the most common results of activity in the comparator circuits is that a specific motor program is labelled 'faulty needs checking', and in consequence is executed more slowly and with greater risk assessment on subsequent occasions. Contextual tagging is not, then, motivationally unbiased—it favours more aversive tags. This aspect of the theory proposed in the first edition receives stronger emphasis in this second edition.

The spatial environment and spatial location constitute the most important forms of context that mammals encounter. Any contextual tagging system, including our originally postulated comparator, must therefore depend critically upon just the kind of spatial mapping function that O'Keefe and Nadel (1978) have attributed to the hippocampus. The converse is also true. The three-dimensional complexity of a spatially extended environment requires for its analysis a system that is capable of determining a detailed contextual specification of the conditions under which different action patterns are correct: if one starts at position P aiming for a goal in location L, action pattern A is correct, but if one starts at position Q, action pattern B is correct, etc., for many different possible starting positions (not to mention many different possible goals in the same general environment). Given our general model, therefore, it is unsurprising that hippocampal function is critical for spatial cognition. We differed, and continue to differ, from O'Keefe and Nadel's position, however, in not restricting context to its spatial variety. As noted above, Davidson and Jarrard's (1993) experiment, showing after hippocampal damage loss of control over behaviour by the completely non-spatial form of context that is hunger, demonstrates that such a restriction would be incorrect.

With regard to memory, here too the role of contextual tagging in the hippocampal contribution is at once apparent. The kind of memory that is 'lost' after damage to the hippocampal formation is precisely that form which, when intact, is tagged for context not only spatially but also temporally and in other ways (e.g. gustatory or olfactory), and so made safe from the effects of interference. In terms of the model of the first edition, this is because the hippocampal comparator system, having completed its work of analysis, sends return messages to the temporal-lobe memory stores that it has been interrogating so as to add updated contextual tags to the relevant stored memories. Destruction of the hippocampus impairs this tagging capacity, so preventing the further memory formation from proceeding correctly but not eliminating memories already stored (and thus giving rise to anterograde but not retrograde amnesia—except to the extent that consolidation is interrupted—as in fact observed in patients with this type of damage).

In a study with marmosets we have recently obtained particularly convincing evidence supporting that part of the above account which postulates that the hippocampus does not itself directly store memories. After excitotoxic damage to the CA1 cell field, the animals could no longer perform a conditional discrimination learned before the lesion or learn new conditional discriminations. After transplantation of a neuroepithelial stem cell line reconstituting the damaged hippocampus (Sinden *et al.* 1997), however, the

marmosets were not only able to solve new conditional discriminations but also successfully retrieved the original discrimination learned prior to the lesion (Virley *et al.* 1999; see also Gray *et al.* 1999). This result clearly demonstrates that the lesion-induced impairment affected the handling of stored information, but left that stored information itself intact.

1.9 A COMMON COMPUTATIONAL CORE FOR SEPTO-HIPPOCAMPAL FUNCTION—MARK 2

Up till this point we have emphasized the developments in the literature that have strengthened our original theory in comparison to both older and more recent competing theories. These developments have provided confirmation with totally new classes of drug of the essential drug–lesion analogy on which the theory is based; confirmation also by these same drugs of the correlation between anxiolytic action in the clinic and changes in the control of theta activity in rats; and demonstration that anxiolytic drugs impair the most memorial of hippocampal-sensitive tasks, and do so even when injected into specific loci in the brain so as to affect solely the control of theta activity. The heart of the theory—that anxiolytic drugs act on a behavioural inhibition system which overlaps substantially at the neural level with the septo-hippocampal system—seems healthier than ever.

However, our original theory consisted not only of a number of fairly global axioms and general predictions derived from them, but also of detailed specification of the neural mechanisms thought to underlie the axioms (very superficially described in the previous section). This specification went quite deliberately beyond the available data in many cases to produce as full and complete a model as we could at the time. It contained (as indeed this edition is also forced to contain) many guesses and assumptions. The bad news is that quite a few of these detailed guesses and assumptions have proved wrong. The good news is in two parts: the guesses and assumptions were couched in such a way that each could be separately proved right or wrong by fairly simple experiments; and, where they have proved wrong, or where completely new data are available, we have been able to propose a simpler and more coherent mechanism for what is, nonetheless, essentially the same overall theory.

Our first modification is not, in fact, driven by new data but by deeper consideration of the old data with which we started the previous section. The anatomy of the hippocampal formation is relatively simple, implying a single fairly basic transformation of its inputs, with the differences in its various apparent functions being determined by the structures receiving its output rather than by its performing a range of truly different functions (Fig. 1.3). Yet, in the first edition, despite the attempt to assign to the hippocampus a consistent form of operation which could combine memory, anxiety, behavioural inhibition, etc., we nonetheless implied that the hippocampus received information about stimuli, memories, responses, motor programs, and plans which it then kept distinct and processed in order to produce its output. In retrospect, even this attempt at simplification seems too complex for such anatomically homogeneous circuitry. We also implied that the hippocampus carries out a wide range of computations—checking,

detecting mismatch, inhibiting, tagging, and so on—and with results that encompass everything from preventing amnesia to generating anxiety. Again these processes seem unduly complex for such simple circuitry. Here we resolve these problems through a linked set of ideas: that all inputs to the hippocampus represent goals (conflating the stimuli, memories, responses, motor programs, and plans of the first edition); that the hippocampus detects when there is conflict between concurrently active goals (conflating the just checking, comparator, prediction, and mismatch functions of the first edition); and that, when conflict is detected, the hippocampus produces output which increases the valence of affectively negative stimuli and associations (conflating the behavioural inhibition, tagging, and memorial functions of the first edition). In addition, we extend the critical circuitry to include the entorhinal cortex and the posterior cingulate cortex, and postulate additional logical gates within the septo-hippocampal system which determine its precise outputs under different conditions. We thus have more circuitry doing what is computationally, if not at the psychological level, a simpler job.

'Goal' as an explanatory concept in psychology has always been contentious (George and Johnson 1985). However, we use it simply in the sense described by Hinde (1982, p. 307; see also Towe and Luschei 1981):

> Some authors have labeled behaviours as 'directive' or 'goal directed' on the sole criterion that variable means are used to achieve a consummatory situation . . . [but if] each type of behaviour is stereotyped, its cessation could be due merely to inhibitory effects consequent upon performances, rather than to error signals. . . . [In contrast] if rats are subjected to spinal or cerebellar operations so as to interfere with their motor coordination, they may nevertheless use quite novel movements to make errorless runs through a maze. The essential point here is that the new movements are not stereotyped, but selected from variable patterns in such a manner as to bring the animal nearer the goal. Furthermore, the new patterns are 'directly and efficiently substituted without any random activity'.

For our current purposes, the most important feature of the term 'goal' is that it can be distinguished from both 'stimulus' and 'response' in their pure sense—and represents a necessary conflation of the two. The goal of the rat running down the runway has both a stimulus (in this case, place) component and a response (the animal's tendency to run towards it) or motivational component. Take either of these aspects away and it would cease to be a goal. As discussed in detail in Chapter 7, we see the firing of hippocampal cells as reflecting information about goals. Goals often have fixed locations in space, hence the place fields often found with hippocampal cells. But they may not, hence the non-spatial fields seen in the majority of cells in non-spatial tasks. Furthermore, they depend on the animal's (as opposed to the experimenter's) intent, hence the changes in the location of place fields when reinforcement contingencies change, and the fact that hippocampal cell firing does not always reflect the contingencies which experimenters think that they have embodied in the paradigm.

This is not to say the septo-hippocampal system codes goals as such. On our model this system receives information from brain areas which code goals and so its activity mirrors the specificity of encoding of those areas. However, it receives this efference copy of the true encoding of the goal information only so that it can detect when more than one goal is concurrently highly activated, i.e. when there is goal conflict. For this purpose

the hippocampus itself need only know that there are a number of goals, not precisely what they are. Indeed, the same hippocampal cell could, in principle, respond to a number of different goals (with more goals producing more activity in the cell in question). However, the area from which it receives the information will, at any one time, tend to code for only one goal, giving the hippocampal cell an apparent selectivity of information coding. (This selectivity derives, in our view, from the fact that motor programming areas are topographically mapped into the septo-hippocampal system, producing a mapping of what is essentially 'goal space'; see Appendix 4.)

Goal information comes from all areas which can be concerned with specifically goal-directed behaviour (including areas as diverse as the amygdala and the cerebellum). Note, however, that the sequencing of goals that is the true 'planning' function is held to take place in the prefrontal cortex, perhaps the most prominent area among those that send to the hippocampus information about goals. In the first edition, the hippocampus was held to receive information about responses. Here, we hold that the hippocampus deals only with conflicts between goals (approach–avoidance conflict in nominal tests of anxiety; S+ approach/S– approach conflict in nominal tests of memory) and not with conflicts between responses or motor acts as such. Thus, in the learning of mirror drawing there is considerable conflict between old and new muscular patterns required to achieve the goal; but there is only one goal and so no goal conflict, and there is motor system rather than hippocampal involvement. That is why someone like H.M., who lacks a hippocampus, can learn mirror drawing as well as anyone else.

When it receives information about only one prepotent goal, the hippocampal formation is, effectively, in 'just checking mode'. When, in addition, it receives information about a second conflicting goal, it detects the multiplicity and effectively enters 'control mode'. Note that no real change in the nature of processing occurs between these two modes. At all times information enters the comparator and is integrated, but when the summation of inputs passes some threshold it also produces output.

The output from the hippocampal formation involves the return of information to those areas whose activity has given rise to conflict. At present we view the hippocampal output as providing a simple, quite general, affectively negative biasing function. The effect is to increase the valence of affectively negative stimuli and associations of stimuli (memories), and hence to shift approach–avoidance conflict in the avoidance direction both on the current and future occasions. The net result of this process is determined, however, not by the hippocampal formation but by its target structures. The details of how this is achieved determine our interpretation of the functional role of theta activity.

The mechanism we postulate relies upon the concept of recursive networks. To introduce this concept we first describe Marr and Poggio's solution to the global stereopsis problem (summarized and illustrated by Frisby 1979, pp. 150–1). We use their model solely as an analogy but it should, nonetheless, give an idea of the power and apparent sophistication which recursion can produce in the processing of information by even simple parallel networks of the type we postulate as linking the hippocampal formation with goal-processing structures.

'Global stereopsis' is the capacity to see depth, and as a result the form of an object, in random dot stereograms presented to the two eyes. The stereograms themselves (Fig. 1.6A) are made by first putting black or white squares into a grid in a totally random

pattern. This is the part of the stereogram to be presented to one eye. The second part of the stereogram is then created by first copying the first part and then moving portions by calculated amounts equivalent to the disparity on the retina which would be produced by an object at an appropriate distance away from the viewer. The original location of the moved items is filled with further randomly chosen dots. As a result, when viewed alone, both halves of the stereo pair are effectively totally random. When the stereograms are viewed through a stereoscope for sufficient time (often 30 seconds to 2 minutes is required), the depth, and hence form, of the object coded in the stereogram become visible (Fig. 1.6D). At this point, every dot in one stereogram has been matched by some mechanism in the brain to a dot in the other, the specific disparity calculated, and depth deduced. Since the dots themselves are indistinguishable one from the other, this is a daunting computational task when there are many thousand such dots.

Figure 1.6C shows a simple network devised by Marr and Poggio (see Frisby 1979, pp. 150–1 for summary) which solves this computational problem. The key features of this network are: (a) the input patterns from the two eyes are 'clamped' onto the input lines of adjacent sides of the square lattice; (b) lateral facilitation and lateral inhibition within the network then change the activity at each of the nodes where the two types of input line intersect; (c) the result of this change is, in turn, fed through the network (recursively) and the process is repeated until the network settles into a stable state. This last step may require a very large number of recursions. If we assume that the human visual system solves the problem in the same way as the network model, and that a single recursion takes, say, 50 ms, then the typical undergraduate in a psychology practical class has a nervous system which requires of the order of 1000 successive iterations to solve the problem. However, because these calculations are done recursively, feeding the output from one calculation back as input for the next, only a few (two or three) layers of neurons are required for all of these calculations for any particular point in space.

Lest this detailed treatment of global stereopsis should tempt you to think that the hippocampus, if it works in the same general way, must perform some kind of visual figure-ground calculation, it should be noted that similar recursive circuitry is used by Hinton *et al.* (1993) for 'clean up' units which adjust discrepancies between words and semantic units in a neural network simulating reading. This matches the global stereopsis model in terms of the fact that recursion is used to resolve differences between conflicting sources of input, but cannot be viewed as a figure-ground computation in any conventional sense. A particularly interesting property of the Hinton *et al.* model is that it has been constructed in such a way that changes in its behaviour when it is 'lesioned' can account for many of the changes resulting from damage to language areas in the human brain.

The septo-hippocampal system is richly supplied with the means to interact recursively with cortical (e.g. Fig. 10.5, p. 262) and subcortical areas, so as to extract higher order information. It has been something of a puzzle that the cortical and subcortical inputs to the hippocampus appear to carry very similar information (Appendix 4). If, however, the septo-hippocampal system carries out calculations akin to those required for global stereopsis, these various sources of information might well be expected to differ only subtly, as do the images from the left and right eye, or word and semantic code. The septo-hippocampal system might, then, be concerned to extract the higher order

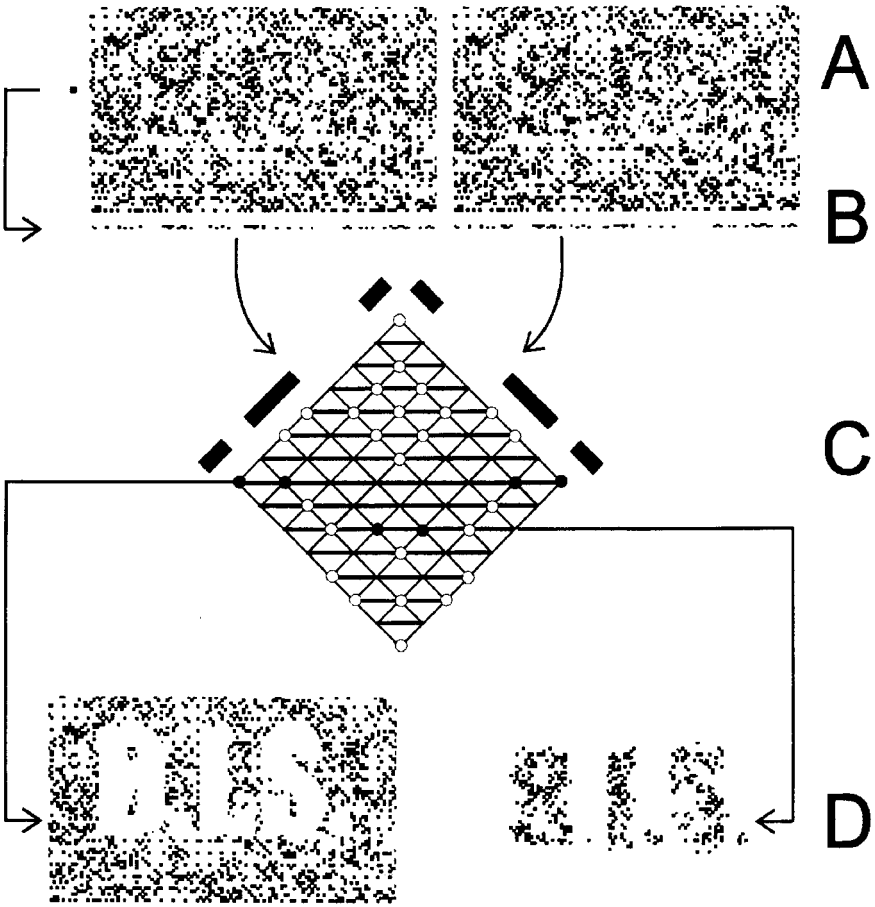


Fig. 1.6 Marr and Poggio's computational solution to the global stereopsis problem (for more detailed description, including simulation outputs, see Frisby 1979). (A) Random dot stereograms are constructed by first creating a purely random image (left-hand side) and then taking part of it, here an area defined by superimposing the letters B.I.S., and shifting the dots a small amount to one side. Dots in the area shifted to are erased and the blank space created by the shift is filled with more random dots. This creates a second image which, by itself constitutes random dots (right-hand side) but which, taken together with the other half of the stereo pair, defines two areas (B.I.S. and its background) which differ in retinal disparity and hence implied depth when viewed binocularly. (B) A slice taken horizontally at the point marked by the filled square at the left of the left-hand image in A. (C) Small portions of B taken from the equivalent areas for the left and right eye and superimposed as input onto a computational lattice. Each intersection in the lattice represents a cell (or node). Thick lines represent reciprocal excitatory connections between cells, thin lines reciprocal inhibitory connections. The imposed input from the left and right eye will initially cause a large number of nodes to be activated by coincident input from the two eyes. This will include not only cases where a dot in one image is correctly matched to a dot in the other (filled circles) but also cases where an incorrect match, or 'false fusion', occurs (open circles). In terms of analogy with our view of the B.I.S., it is the presence of these incorrect matches that is the equivalent of conflict. The excitatory and inhibitory connections within the lattice then, over a large number of cycles, eliminate activity in the nodes representing incorrect matches,

basis of the differences (the equivalent of depth for stereograms, or of meaning for the case of reading words). Given the complexity of the circuits in which, as we shall see later in the book, the hippocampus is embedded, we should be prepared for the septo-hippocampal system to be capable of dealing with many different higher order properties of the same input information.

Given the power of recursive networks, what use might the hippocampal system make of them? Our general answer to this question is that the hippocampal system acts so as to increase the valence of adverse outcomes associated with items (locations in space, associative bonds in memory, or motor programs). This 'association' can be innate or acquired. In relation to memory, this general approach makes our position the antithesis of most others (including that of Eichenbaum *et al.* 1994, with whose views we otherwise have much in common) by supposing that the hippocampal formation acts, not to strengthen associations that are retained, but to weaken those that are discarded. Consistent with this position, there is evidence that the errors made in memory tests by patients with hippocampal damage over-represent items that would have been correct under slightly different circumstances. Such a pattern of errors cannot come from enhanced forgetting, as would be predicted if the role of the hippocampus were simply to facilitate the formation of (correct) memories. It *is* predicted, however, if this role is to eliminate competing (incorrect) memories. Similarly, our theory predicts that hippocampal deficits will be seen only in those tasks in which interference is actually overcome by control subjects, not in those in which interference is present but not overcome by controls.

In summary, then, the common computational core postulated in our model as lying at the heart of hippocampal function, whether this is manifest in spatial cognition, memory, or anxiety, comprises a comparator, to detect conflict between concurrent activation of incompatible goal-seeking mechanisms (and hence mismatch and/or threat in the language of the first edition), coupled to recursive networks that link the septo-hippocampal system to memory stores (in the temporal lobes), complex motor programming circuits (in the basal ganglia and frontal cortex), 'fixed action pattern' circuits (in, for example, the amygdala and hypothalamus), and thalamocortical perceptual systems, with all of which the septo-hippocampal system interacts so as to increase the weighting of items with greater potentially adverse significance (Fig. 1.7).

The septo-hippocampal system produces such an increase in negative weighting at every cycle of a series of computations which progressively increase bias (and hence

---

leaving the activity representing correct matches. Note that it is not the initial level of activity in any node that determines whether it will remain—rather it is the relation between it and the recursive changes in activity within the network. (D) Each horizontal line of nodes in C represents a particular retinal disparity and hence depth. Once activity in the network has stabilized, only the correct matches (filled circles) remain active. Their output is then sent to different arrays, each topographically mapped with respect to the visual world, representing particular depths (see arrows connecting C to D). These have the 'ground' (left) and 'figure' (right) separated one from the other. In terms of analogy with our view of the B.I.S., it is the reduction of a very large number of conflicting possibilities (represented by all the matches, correct and incorrect, in C) to the single possibility represented by the right-hand image in D that is the same general class of operation as that carried out by the B.I.S.

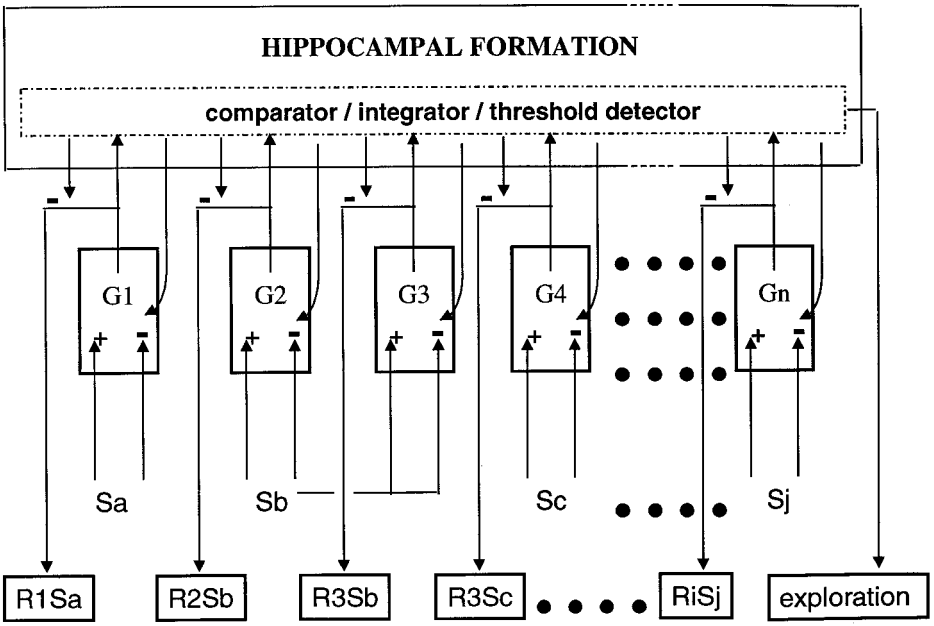


Fig. 1.7 The hippocampal formation as a general device for the resolution of goal conflict. A goal is defined as a particular combination of environmental stimulus (S) and response tendency (R). Often, two goals will differ in both of these attributes (R1Sa versus R2Sb). However, some stimuli may elicit more than one response tendency (R2Sb versus R3Sb), e.g. a goal box which elicits both approach (because of a prior association with food) and avoidance (because of a prior association with shock). Likewise, some responses can be addressed to more than one stimulus (R3Sb versus R3Sc) as in a lever press response which may be elicited by either a left or a right lever (or both) depending on the prior history of reinforcement. Any stimulus (e.g. Sa) is held to activate a particular goal (G1) defined by the conjunction of that stimulus and a particular response (R1Sa). The strength of activation is determined by integration of the prior affectively positive (+) and affectively negative (–) associations of that goal. In many cases the goal (say G1) will be the only one with sufficient net activation to produce a response; or, if another goal is activated, it will nonetheless be the clear winner. In this case the response (R1) is directed to the relevant stimulus (Sa). The hippocampal formation will receive an efference copy of the output from G1, but in this case will not react. Conflict is defined as the case when two or more goals (e.g. G1 and G2) are nearly equally and highly activated. The hippocampus as a comparator determines the degree of conflict and the greater this is, the greater will be its output. The output has two effects. First, it blocks the output from the areas defining the goals (e.g. G1) to the areas responsible for achieving them (R1Sa). Second, it increases the value of any affectively negative associations (including, in humans, inferences) specific to that goal. In many cases this will allow resolution of the conflict in favour of that goal which has the lesser negative associations. Output from the hippocampus is increased, recursively (as in the global stereopsis model of Fig. 1.6), until conflict is resolved while, at the same time, commanding exploratory/risk assessment activity which can provide additional information (either devaluing existing negative information or providing new negative information) which allows the conflict to be resolved.

suppress goals) until one or other goal is sufficiently predominant to take control of the motor mechanisms. This increase in negative bias has two consequences. It affects current motor output directly, and it affects future motor output indirectly through its biasing of associations (including those that are just being formed). Additionally, while its recursive operations continue to fail to resolve the conflict, the septo-hippocampal system produces additional output which engages exploratory mechanisms designed to resolve the conflict through the obtaining of new information about the environment (thus preventing the animal from getting locked into an impasse). We also presume that during consolidation, and particularly during sleep (Appendix 7), the hippocampus can direct a similar 'internal exploration', which in human beings may best be thought of as rumination.

In Marr and Poggio's model, since it is carried out on a digital computer, each cycle of computation is quite discrete. However, the effects of opening ion channels on hippocampal neurons can change neuronal excitability for periods of time which equal or exceed the time it takes information to circulate from the hippocampus to its targets and back (Miller 1991, 1995). The effect of theta activity (and particularly the imposition of a regular period of inhibition on hippocampal cells by the septal input) will be to quantize hippocampal processing and thus reduce the chances that the effects of information received on a previous cycle interfere with the comparator's assessment of inputs on the current cycle. From this follows the conclusion that theta activity will be important, particularly for complex or evenly balanced conflicts; but, since it only increases the acuity of processing in the temporal domain, theta activity will not be essential for the hippocampus to perform at least some functions where a mere degradation in performance is not crucial. This view that theta activity merely enhances rather than being absolutely essential to hippocampal function contrasts with the more absolute position taken both in the first edition and by O'Keefe and Nadel (1978).

Here we should note also that the extensive circuitry controlling the frequency of theta activity and gating its occurrence implies that particular levels of that activity will not be universally beneficial (or theta would occur all the time at a fixed level). Nor, therefore, should we assume that the interaction of the hippocampus with other brain structures will always be beneficial. Certainly, there are some behavioural tasks (e.g. two-way active avoidance in a shuttlebox) where lesions of the septo-hippocampal system produce an improvement in performance.

A final point to note is that goal conflict may result from the activation of quite different response systems (e.g. approach, avoidance) or from the activation of an essentially single response system (e.g. lever pressing) equally by two different stimulus sources (left lever, right lever). While Fig. 1.7 shows a single circuit plan which can resolve both types of conflict, the full version of our theory (Chapter 10) ascribes the resolution of what can be thought of as response–response, stimulus–stimulus, and response–stimulus versions of goal conflict to different parts of the overall septo-hippocampal system.

1.10 THE ROLE OF OTHER BRAIN REGIONS

The brain is a highly interconnected system. It is not surprising therefore that we find it impossible to speak of the septo-hippocampal system without repeated reference to

the many other parts of the brain to which it is closely linked. Indeed, we see the apparent functions of the septo-hippocampal system as arising from the additional properties it gives to a highly distributed network in which key aspects of information processing depend on the properties of its targets. This will already be apparent from the foregoing pages, and it will become still more so as the book unfolds. In this introductory overview we plant just a few signposts.

Figure 1.8 expands on the simple defence hierarchy of Fig. 1.2 and displays a coherent neurology, borrowed largely from LeDoux (1994) and Graeff (1994), setting out much of what is known about the brain's defensive systems. This neurology starts with the dorsal periaqueductal grey as a key centre for the control of immediate responses to a predator: freezing, fight, flight, autonomic output, and analgesia. Specific lower subsystems controlling separate ones of these different reactions are also known, but differentiation

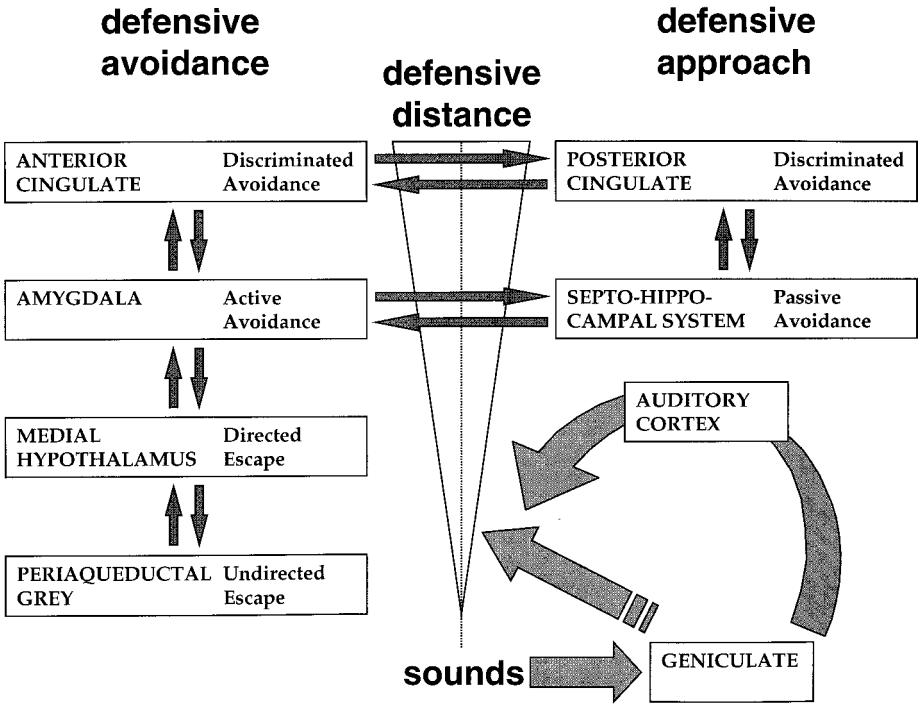


Fig. 1.8 The defence system viewed as a hierarchy of structures. The higher levels are engaged by increasing defensive distance and there are two parallel streams which control behaviour when danger is to be avoided or approached, respectively. Both defensive distance and defensive direction (approach/avoid) are determined by input from sensory systems which are also hierarchically organized. A simplified diagram of the auditory system is given as an example. Following LeDoux (e.g. 1994) we see this as consisting of a relatively simple, direct ('quick and dirty') route which passes directly from the geniculate to areas such as the amygdala and a relatively complex, indirect ('slow and sophisticated') route via the auditory cortex. In the same way that activity in the higher levels of the defence hierarchy can override the behaviour which would result from activation of the lower levels alone, more sophisticated information can override the effects which would be produced by 'quick and dirty' stimulus information (albeit with a slight time lag).

between them is not necessary for our present purposes. The periaqueductal grey is activated by an immediate predator, by an immediate dominant conspecific, by pain, and by high levels of carbon dioxide. As noted earlier, Graeff (1994) has suggested that activation of the periaqueductal grey is the basis of panic attacks. The flight (escape) elicited by stimulation of this region is undirected, so much so that the animal is likely to crash into obstacles rather than leap over them.

The dorsal periaqueductal grey is connected to the medial hypothalamus, where more sophisticated escape mechanisms appear to be located. This in turn is connected to the amygdala, which coordinates simple avoidance. In LeDoux's formulation, the amygdala is then viewed as being connected to the hippocampus, which mediates complex, particularly inhibitory, avoidance. However, in Chapter 6 and Appendix 3 we find extensive reasons to see the anterior cingulate as representing several additional levels of the active defence system, and hence as lying directly above the amygdala in the 'danger avoidance' hierarchy. By contrast, the hippocampus proper and, above it, the posterior cingulate appear to be parts of a parallel 'danger approach' hierarchy.

Figure 1.8 also shows, at bottom right, multiple levels of one of the hierarchically organized sensory systems—audition. As with the output systems, each level of sensory processing deals with more highly processed items. As discussed in Chapter 6, several of these levels are connected directly to the amygdala, as are equivalent levels of other sensory systems. (The main cortical area which does not project to the amygdala is the parietal cortex.) This leads to a very neat account (LeDoux 1994) of the functioning of the avoidance system.

Activity at each level of each sensory system is passed to the amygdala directly and in parallel. Note that this is not redundant as, for example, the geniculate codes for a simple CS, while the auditory cortex codes for the discrimination between a CS+ (associated with punishment) and a CS– (associated with the absence of punishment), and so on. In each case, then, a particular type of avoidance response can be produced simply by strengthening (via long-term potentiation; see Appendix 2) the appropriate connection in the amygdala. This can link directly, for example, neurons which code for a simple auditory stimulus in the geniculate to neurons which control avoidance motor programs.

An important corollary of LeDoux's formulation is that each of the nuclei in the hierarchical active defensive system (from the periaqueductal grey to the anterior cingulate) must be presumed to be receiving sensory information at all times. Hence, given adequate stimuli, all of them will be concurrently pre-programming a range of quite independent potential responses. This multiplicity of potential responses may seem complicated but it can easily be controlled by a system which depends on release (inhibition of inhibition), rather than activation, to control responses (see, for example, McNaughton 1989, Chapter 2). The concurrent high-intensity activation of multiple networks, coupled with control by release from inhibition, allows the sudden dramatic shifts which can be seen in experimental animals between, say, freezing and defensive attack. This can be attributed to a shift of disinhibitory influences from, for example, a (continuously updated) freezing motor program to a (continuously updated) defensive attack motor program.

As with modern network models of many processes, all of these structures are reciprocally connected. Despite our description of, for example, the dorsal periaqueductal grey

as controlling undirected escape behaviour, it is probably much better to view this structure as part of a distributed system, including the hypothalamus, amygdala, anterior cingulate, posterior cingulate, and hippocampus, in which certain patterns of activity engage the various fight–flight–freeze nuclei and include them within the effective distributed network. The periaqueductal grey may then be the most important node through which information passes during fight, flight, and freezing, but it would be an error to think of it as the only structure in the defensive system active at that time, or even as the only structure contributing to escape.

Our analysis of the defensive system has, so far, closely followed those of LeDoux and Graeff (see Chapter 6 and Appendix 2). However, we differ from them in thinking that the hippocampus should not be viewed as, say, a source of contextual stimuli which can come to control the amygdala (in which case there should be strong hippocampal–amygdala connections rather than, as is predominantly the case, the other way round). Rather, we think of the hippocampus as a device which can resolve conflicts (as outlined above), not only between separate programs concurrently activated within the amygdala, but also between programs in the amygdala and those outside the defence system. It is important in this context, as with the case of mirror drawing, to emphasize that the septo-hippocampal system is involved in the resolution of goal rather than motor conflict. In LeDoux's and Graeff's formulations the lower levels of the defence system (both for stimulus processing and response production) can be overridden by the higher levels. Thus, if a sophisticated avoidance response is being programmed, this prevents the occurrence of counterproductive undirected escape. Resolution of this type of conflict between levels is solely the business of the active defence system (via, for example, inhibition of the periaqueductal grey by the amygdala). However, where there are two equally attractive (or equally unattractive) alternatives to which to direct incompatible avoidance responses, then interaction between the hippocampus and the amygdala (or other related areas) will be required to resolve the conflict between the different goals.

The defence system is, of course, only one (rather specialized) system controlling goal-oriented behaviour. At the time of writing the first edition of this book, how the septo-hippocampal system was able to influence more general motor control areas was a considerable enigma. A solution to this problem was forthcoming with the demonstration (Kelley and Domesick 1982) of a major projection from the subiculum (the main output station from the hippocampal formation) to the nucleus accumbens (ventral striatum), described by Mogenson and Nielsen (1984) as the gateway to the motor programming circuits of the basal ganglia. Onward transmission from the nucleus accumbens goes by two routes: downstream via the substantia nigra to affect ascending inputs to the caudate-putamen, and upstream via the ventral pallidum and dorsomedial thalamic nucleus to affect the prefrontal and cingulate cortices. It is almost certainly via the subiculo-accumbens projection that the hippocampal formation is able to interrupt many prepotent motor programs (Gray *et al.* 1991). Still more recently, it has become clear that the subiculo-accumbens projection provides, in addition, a route by which hippocampal output can influence perceptual function. This can be achieved by way of the projection, described by Lavin and Grace (1994), from the nucleus accumbens, via the ventral pallidum and nucleus reticularis thalami, to virtually the entire thalamocortical sensory processing system. Analysis of this projection (Gray *et al.* 1997) suggests

that activation of this subiculo-accumbens–nucleus reticularis pathway should provide an overall boost, by disinhibition, to sensory processing. Thus, two of the key outputs of the behavioural inhibition system (Fig. 1.1), interruption of ongoing motor programs and increased attention to the perceptual world, are perhaps achieved simultaneously by activating the subiculo-accumbens projection.

We have mentioned above the prefrontal and cingulate cortex (the sites, it will be recalled, at which psychosurgery is practised for the treatment of certain drug-resistant forms of anxiety) quite casually. Encompassing, as they do, a huge portion of the cortical mantle, our view of their precise function and organization (and hence their interaction with the septo-hippocampal system) will require an entire Appendix (3; and see Chapter 6) for its exposition. Briefly, we view them (following, for example, Barbas and Pandya 1991; Fig. 1.9) as being hierarchically organized areas which deal (in their successively 'higher' layers) with progressively higher levels of anticipation of action. Thus the frontal eye fields can be viewed as pre-pre-motor cortex and area 46 (with its role in working memory), as pre-pre-pre-motor cortex. In the same way, then, that we distinguished the role of the hippocampus (in resolving concurrent goal–goal conflict) from the role of the defence system and other motor systems in resolving motor program conflicts without goal conflict, so we must distinguish its role from that of prefrontal and cingulate cortex.

(a)

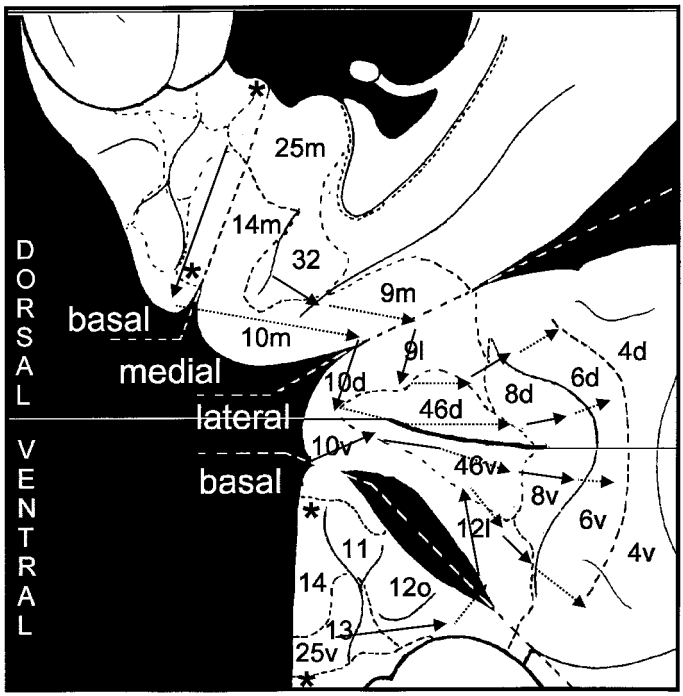


Fig. 1.9 A simple hierarchical view of frontal cortex. (a) The medial, lateral, and basal aspects of frontal cortex partially unfolded. (Basal is repeated with a star marking points at which the medial and basal aspects of areas 14 and 25 are joined.) The arrows move from less laminated and myelinated cortex to progressively more laminated and myelinated cortex (b) See next page.

(b)

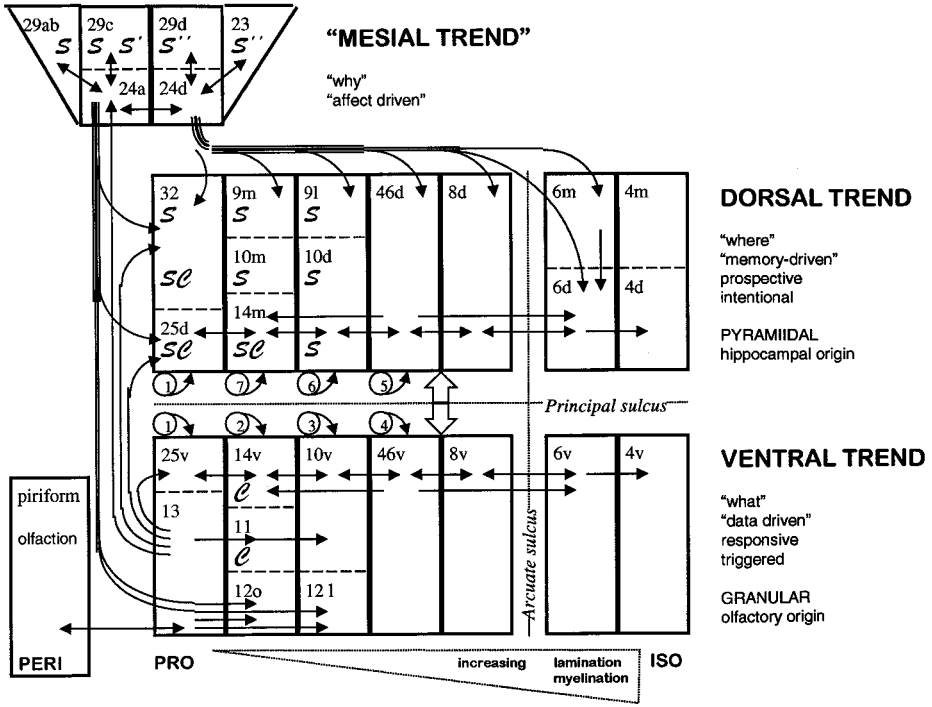


Fig. 1.9 A simple hierarchical view of frontal cortex. (b) Frontal and cingulate cortex fully unfolded to show their organization. (Areas of boxes do not reflect actual amount of cortex.) The cortex is proposed to be organized in terms of three separate 'trends': dorsal, ventral, and mesial (i.e. cingulate cortex). We tentatively suggest that the dorsal and ventral trends can be viewed to some extent as mirror images of each other and that the progression of areas from 4m through 32/25d and from 4v through 25v/13 represent a steady increase in the level of essentially the same kind of processing (see text). *S* or *C* inserted into boxes shows input from subiculum or CA1 of the hippocampus, respectively. The numerals 1–7 in small circles represent the topographically organized input from the thalamus, each segment of which innervates an entire level of dorsal or ventral trend. PERI, periallocortex; PRO, proiscortex; ISO, isocortex.

In our view these cortical areas are involved, quite independently of the hippocampus, in the resolution (i.e. ordering) of conflicts between successive sub-goals in a task. In the case of prefrontal cortex this amounts to saying that it is concerned with plans more than goals as such. However, where (as is common in certain types of working memory task) there is concurrent goal conflict within such a task, both the septo-hippocampal system and the prefrontal cortex are likely to be involved.

Our clinical neuropsychology of anxiety, and particularly of the anxiety disorders, is described in detail in Chapter 11 but is briefly summarized in Fig. 1.10. It follows very closely from the hierarchical view of the defence system obtained from animal experiments. A particular symptom (e.g. panic or obsession) is deemed to result from a high level of activity in the relevant site indicated in the diagram. However, this symptom may result

from perfectly normal activity in the defence system in response to appropriate (usually extreme) environmental stimuli; or from pathological (biochemical or neural) activity in the site concerned; or (which gives rise to most of the problems of differential diagnosis in the clinic) from pathology elsewhere in the system, since this, as indicated, has multiple reciprocal connections. Worse still, our theory holds that a prevalence of, say, panic symptoms does not necessarily indicate a disorder of the structures that (proximately) mediate these symptoms, as they may well be a secondary consequence of disorder in some other part of the defence system. Our concentration in this part of the argument on identifying specific brain regions involved in the specific symptoms and syndromes of anxiety leads us to emphasize, in Chapter 11, the differences in function between each such region. This emphasis is redressed in the final two chapters, dealing respectively with the personality traits (and especially their heritability) that predispose towards the anxiety disorders, and with the effectiveness of psychological treatments of these disorders. Data from both these areas of research provide striking evidence that, despite the complexity of the brain systems which underlie the phenomena of anxiety and the anxiety disorders, there is an underlying unity that binds them together. It is this unity that continues to justify the title of our book: the neuropsychology of anxiety.

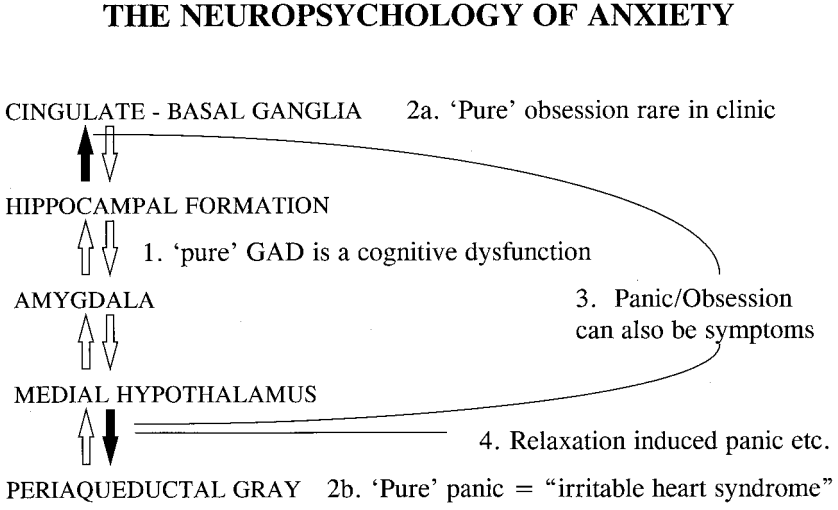


Fig. 1.10 A summary of the key features of the proposed neuropsychology of anxiety. Panic is directly controlled by the lowest level, the periaqueductal grey; phobia (and the autonomic aspects of anxiety) by the medial hypothalamus and amygdala; the primary cognitive aspects of anxiety by the septo-hippocampal system; and obsession and compulsion by the cingulate and its interactions with the basal ganglia. The implications are that pure generalized anxiety disorder (GAD) may be a primarily cognitive dysfunction with changes in arousal, passive avoidance, etc., consequent on connections between the hippocampal formation and amygdala (1); that pure panic and pure obsession may be the result of dysfunction of the periaqueductal grey and cingulate, respectively (2a, 2b), which will only give rise to anxiety disorder in persons with a neurotic personality type; but that high levels of anxiety, resulting from hippocampal dysfunction, can also produce panic and obsession as secondary symptoms by a spilling over of excitation (3, filled arrows); while, at lower intensity, anxiety can be inhibitory of panic (4).

1.11 CONCLUSION

The purpose of this introductory chapter has been to lay out in broad terms the main themes of the book and the main lines of the theory it contains. We have not dismissed alternative theories of anxiety or the functions of the septo-hippocampal system without careful consideration of their merits and of the evidence in their support. At many points, indeed, rather than dismissing them, we subsume modified versions of these theories into our own model; and in all cases we attempt to incorporate within its explanatory scope the evidence upon which they rest. These efforts to grapple with a wide range of data from many sources and bearing on many specific issues have necessarily resulted in a compendious volume. Thus, besides the topics adumbrated above, the chapters that follow will review still further areas of research that have as yet barely been mentioned, but which are all germane to the arguments we pursue.

The immediately following chapters represent, by and large, only a modest expansion of the contents of the present chapter. This is intended to allow the reader to get a fuller picture of the issues raised here. However, in many cases our interpretations of the data run counter to orthodoxy and depend on quite detailed reviews and analysis. These reviews are cross-referenced in the main body of the book and are provided as appendices in computer-readable[1] form for those who require the details on which our conclusions are based. Even in the appendices, we have not tried to be fully comprehensive and have cited reviews rather than primary sources wherever possible. Despite this, the reader may well feel from time to time, as the authors certainly did, that the wood is getting lost for the trees. It may be useful at such moments to return to this chapter to check our final destination: a theory that simultaneously provides an account of the neuropsychology of anxiety and of the cognitive functions of the various brain systems (most centrally the septo-hippocampal system) that jointly contribute to that neuropsychology. In a summary that the rest of the book seeks to justify: we see the septo-hippocampal system, at the neural level, as being of critical importance to a behavioural inhibition system at the psychological level; and we see these systems as giving rise to pathological anxiety when hyperactive, and to amnesia when hypoactive.

1. The appendices can be found at http://www.oup.co.uk/neuropsych_anxiety

2 Ethology and anxiety

In the first edition of this book, the adequate stimuli for anxiety (i.e. inputs to the behavioural inhibition system) were derived largely empirically, no unifying principle being provided from which they could all be deduced. Recent ethological and ethopharmacological analysis has provided just such a principle. Indeed, analysis of this kind can be thought of as a natural starting point for anyone interested in defining anxiety. Thus, Blanchard and Blanchard (1990a; see also Ursin 1985; McNaughton 1989a) argue that the functional significance of behaviour attributed to anxiety (or other emotions) needs to be taken into account; and that this functional significance 'reflects the dynamics of that behaviour in interaction with the ecological systems in which the species has evolved, implying that these dynamics . . . can be determined far more efficiently when the behaviour is studied under conditions typical of life for the particular species' (Blanchard and Blanchard 1990a, p. 125). Functional analysis of this type can transcend species boundaries, since species-specific behaviours usually have species-general functions and can be characterized by the latter (McNaughton 1989a).

The main problem for ethological analysis is to retain sufficient of the properties of the natural environment to encompass the full gamut of a species' defence reactions, while adding sufficient control of the situation to allow parametric manipulation of stimuli and measurement of responses. The Blanchards have therefore spent many years using 'ethoexperimental' methods to analyse the reactions of prey to natural predators: of rats to cats; of wild rats to human beings; and of mice to rats. The key aspect of their procedures, separating them from the operant procedures described in the next chapter, is their concentration on innate responses to stimuli which the particular experimental subject (as opposed to species) has not previously encountered. An important feature of their work is that, unlike pure field ethology, their experimental methods (and resultant capacity to test the effects of drugs) provide a means of mapping their results with innate responses onto other experimental results obtained with learned responses.

It should be noted that the systems which subserve such innate reactions need not always operate in an immediately 'functional' or 'adaptive' fashion. Instead, they produce responses which have on balance been advantageous in the phylogenetic past but which may not have been advantageous in all situations then, and which need not be advantageous in any situation at all now. This point will be important when we come to explaining how psychological dysfunction can occur in the absence of pathology, and why anxiety disorders are so prevalent. We have argued elsewhere (McNaughton 1989a) that this historical functionality ('teleonomy') should be used to define and hence distinguish all emotion systems. However, in what follows, we simplify with the working assumptions that the functional aspects of a threatening situation are both intuitively obvious to us and currently relevant to the animal. (Ideally both these assumptions should be subjected to rigorous proof.)

It is fundamental to a functional analysis of behaviour that we should be able to compare across species. Patterns of defensive behaviour vary depending on the specific situation, the type of threat, the animal's past history, and many other factors. However, ethologists have managed to categorize this amazing variety of behaviour into a number of fairly simple functional classes. Thus, 'avoidance, flight and defensive responses are found in all animal Phyla' (Eibl-Eibesfeldt and Sütterlin 1990, p. 381) and the functional relationship between these responses is also predictable. Animals avoid when they can, flee if avoidance has not been successful, and defend themselves, usually aggressively, when flight is impossible or difficult. 'The basic principles . . . for birds and nonhuman mammals hold true for man except that man has many cultural expressions of agonal behaviour including elaborations of group aggression in war' (Eibl-Eibesfeldt and Sütterlin 1990, p. 388). These cultural refinements, however, do not require any major theoretical changes. They overlie a foundation of innate reactions, many of which are not only homologous to the reactions of other mammals but also surprisingly similar to them in phenomenological detail. Further, culturally transmitted behaviour patterns are themselves subject to the same adaptive constraints and requirements as genetically transmitted patterns. The similarities between man and other primates are particularly obvious; and human cultural refinements, while extensive, are in any case no different in principle from the learned modifications to agonal (defensive and aggressive) behaviour which occur in other mammals.

2.1 ETHOEXPERIMENTAL ANALYSIS

A problem with classical ethological analysis in the wild is the difficulty of controlling circumstances in such a way as to allow deductions about the underlying control of behaviour. To circumvent this problem, the Blanchards created an apparatus which provided a half-way house between a strictly natural ecology and an entirely artificial experimental environment. This was the 'Visible Burrow System' (Fig. 2.1; Blanchard and Blanchard 1989; see also Blanchard *et al.* 1995 for a mouse version of this test), consisting of an arena (in which the rats could obtain food and water and in which a cat could be presented), connected to which was an artificial burrow system made of plexiglass tubes (in which the rats could live and into which they could escape from the arena). An important feature was that animals could be observed in the burrow system via red or infrared light, which the rats treated as darkness. The rats were established as colonies living permanently in this apparatus and then a cat was briefly presented. This allowed many different types of behaviour to be categorized and the context of the behaviours to be determined through ethoexperimental analysis.

The Visible Burrow System allowed, in particular, comparison of the immediate responses to the predator (as the cat was placed in the arena among the rats) with the pattern of responses after removal of the cat when, from the rat's point of view, the cat was potentially rather than actually present. The relevance of this comparison lay in the fact that 'while such potential threat situations necessarily elicit some defensive behaviours, . . . defensive behaviours such as flight and defensive threat/defensive attack, which work well to present, discrete, threat stimuli, may be useless or even counterproductive

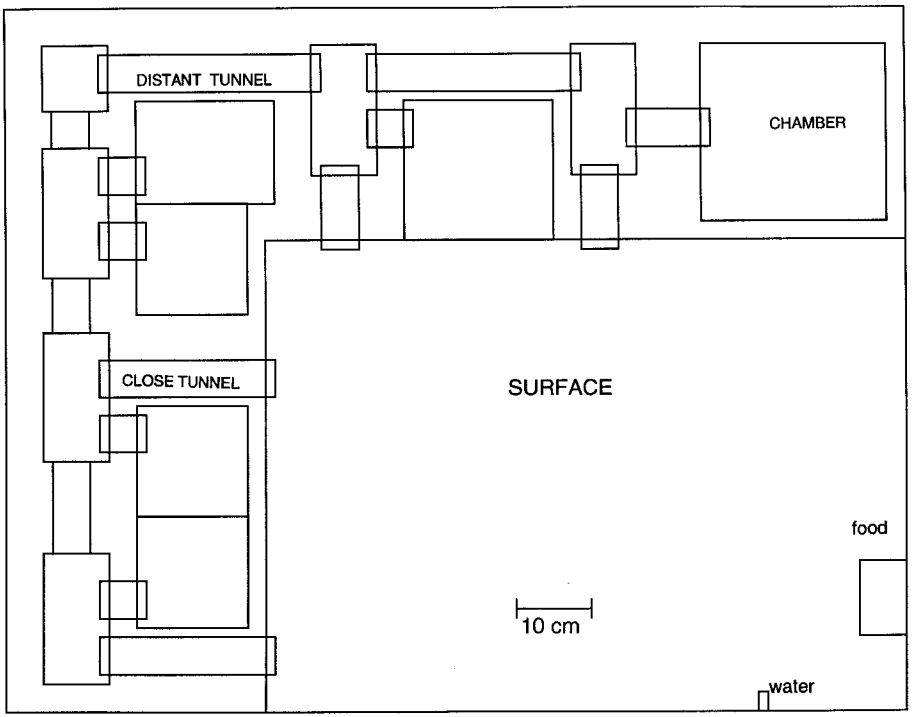


Fig. 2.1 The visible burrow system. (Redrawn from Blanchard and Blanchard 1990a.) For explanation, see text.

in threatening situations in which the actual threatening animal or object has not been localized' (Blanchard and Blanchard 1990a, p. 128).

Using the Visible Burrow System, the Blanchards showed that presentation of the cat produced a characteristic sequence of behaviour patterns. This started with the expected defensive fight–flight responses. Flight resulted, of course, in return to the burrows. There was then behavioural inhibition and freezing within the burrow system, which continued after the cat was removed. As the time since presentation of the cat increased, the rats began to approach the open arena, initially poking their heads out of the burrow. Even once the rats were active and the arena was being entered, behavioural inhibition continued with respect to non-defensive behaviours, such as eating. The pattern observed after removal of the cat could also be elicited simply with cat odour, but here the initial immobility phase was short. On the basis of the strong similarity between the behaviour patterns elicited by recent presentation of a cat and by presentation of cat odour, coupled with the clear common functional aspects of these two situations, they could be grouped together as reactions to a 'potential cat'. A key characteristic of the active reaction to a potential cat was what the Blanchards termed 'risk assessment' (periodic tentative approach, rearing and scanning, and characteristic body postures including thigmotaxis). With longer times after the cat's presence, risk assessment behaviour decreased, and behaviour in general returned slowly to the pre-cat patterns.

To allow parametric and pharmacological analysis of this plethora of defensive behaviour patterns, the Blanchards constructed an ethoexperimental test battery, i.e. a set of limited laboratory situations which allowed a high degree of control over stimuli, but which also used 'a wide enough range of conditions to make it possible for a variety of defensive behaviours to appear and be recognized, and moreover, to suggest by the conditions in which they appeared something of their specific functional significance' (Blanchard and Blanchard 1990a, p. 125). This battery consisted of two separate sets of tests termed, respectively, 'the Fear/Defence Test Battery', which investigated immediate reactions to an actual, proximal predator, and 'the Anxiety/Defence Test Battery', which investigated reactions to a potential predator. A critical feature of the Fear/Defence Test Battery was the use of wild rats which treat humans as predators (this behaviour has been largely bred out of laboratory rats). The behaviour of the predator (i.e. the experimenter) could then be controlled in a way that is impossible with wild predators (e.g. Blanchard *et al.* 1986). The experiments varied both the distance of the human predator from the rat and also whether escape was available or not (for details see Blanchard and Blanchard 1990a), with results summarized in Fig. 2.2. At near-zero predator–prey distances, defensive attack (including vocalization, jumping, and biting) is elicited. At larger distances flight is elicited, if this is possible, or freezing if flight is not possible. The intensity of both flight and freezing decreases with increasing defensive distance. It should be noted that, with the exception of freezing, the behaviour that the Blanchards observed would be inappropriate if the predator were not clearly present and, in many cases, close. The 'Anxiety/Defence Test Battery' employed the odour of a cat or brief presentations of a non-contacting cat as the eliciting stimulus in a battery of tests which allowed measurement of freezing, risk

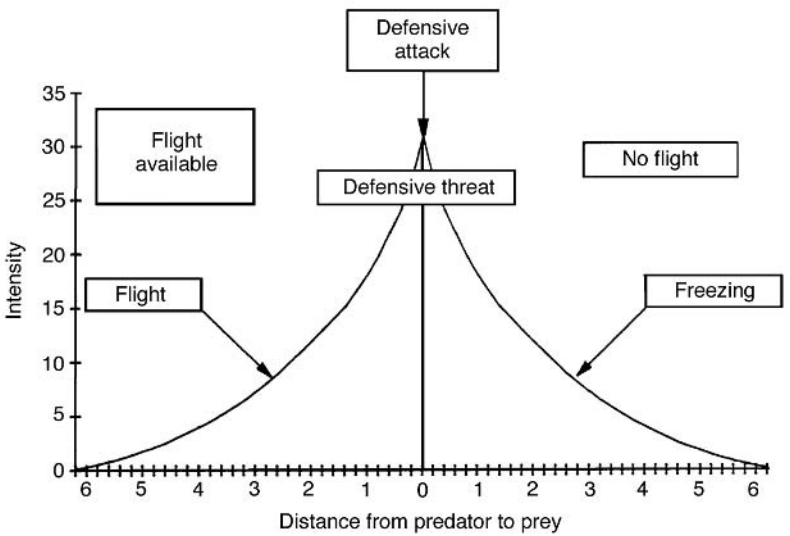


Fig. 2.2 Defensive distance. (From Blanchard and Blanchard 1990a.) As distance from a predator decreases, the intensity of fear is held to increase in an accelerating fashion. This intensity controls very tightly a progression from flight or freezing (depending on whether flight is available) through defensive threat to defensive attack.

assessment, suppression of eating, and other behaviour previously characterized as occurring after exposure to the cat in the visible burrow system.

The Blanchards summarize their overall conclusions as follows: 'natural defensive behaviours of nonhuman mammals are very different in situations involving potential danger as opposed to present danger, and . . . specific behaviours seen in the former situation may be especially relevant to an analysis of anxiety. These behaviours centre around a 'risk assessment' pattern *which includes approach and scanning of potentially dangerous stimuli or situations*, accompanied by changes in posture and movement characteristics. In contrast, reactions to specific, present, threat stimuli involve flight, freezing, and defensive threat and attack. We suggest that risk assessment is the central component of an anxiety pattern, while the reactions to present threat are best characterized as indicating fear' (Blanchard and Blanchard 1990a, p. 124, our italics).

2.2 ETHOPHARMACOLOGICAL ANALYSIS

The effects of anxiolytic drugs also generally support the independent ethological distinction made by the Blanchards between actual and potential threat. In their Fear/Defence Test Battery, anxiolytic drugs were without effect on avoidance of the predator, flight, freezing, or biting (however, they did reduce vocalization and defensive threat). By contrast, in the Anxiety/Defence Test Battery, anxiolytic drugs had a relatively specific effect on risk assessment measures (Blanchard and Blanchard 1990a; Blanchard *et al.* 1993) and, again, did not affect freezing. (It is possible, however, that *conditioned* freezing is sensitive to anxiolytics, e.g. Beck and Fibiger 1995.)

There is a possible parallel here with human clinical phobic reactions. If simple phobia is viewed as a type of anxiety, the failure of anxiolytics to reduce avoidance of, for example, snakes (Sartory *et al.* 1990) is difficult to understand. However, on the Blanchards' analysis, simple object (e.g. snake) phobia is better viewed as the result of fear rather than anxiety, since it involves active avoidance of a present, distinctly localized, predator. Given the Blanchards' results with anxiolytic drugs, we would predict that simple phobic avoidance should not be affected, as indeed it was not in the Sartory *et al.* experiment. The sensitivity to anxiolytics of verbal report (as opposed to physical avoidance) in human snake phobics (Sartory *et al.* 1990) might be accounted for, *post hoc*, if verbal report depends on anticipation of the requirement to approach the snake and, therefore, reflects risk assessment rather than avoidance of the snake itself. The therapeutic situation is clearly one where the distinction between fear and anxiety is difficult to make. For example, the presentation of the snake by the experimenter (unlike the sudden appearance of a snake in the wild) could result in its being perceived as a potential rather than an actual threat, since the patient might be aware that the possibility of being bitten during therapy should have been excluded by the ethics committee before they approved the therapy! Furthermore, presentation of the snake by the experimenter could well create pressure for the patient to approach the snake, and approach to a potential source of danger is a prime characteristic of the Anxiety/Defence Test Battery. These issues would repay further experimental attention (the experiment by Sartory *et al.* has not, so far as we know, been followed up by others).

The specific nature of the drug effects on risk assessment measures which the Blanchards observed is also interesting and emphasizes that care must be taken with parametric aspects of tasks if they are to be interpreted properly. Across different components of their anxiety/defence battery, the Blanchards found that anxiolytic drugs both increased and decreased risk assessment measures. As the time since presentation of an actual cat increased, the control rats in the visible burrow system went from freezing to high levels of risk assessment and then, finally, to low levels of risk assessment and a return of normal daily routines. In specific tests of the anxiety battery in which an actual cat was used to generate anxiety, risk assessment was low in controls and was increased by an anxiolytic drug. In tests where cat odour rather than an actual cat was used, risk assessment was high in controls and was reduced by the drug. Thus, in all cases the drug-induced change in risk assessment could be interpreted as an increase in effective defensive distance and hence a decrease in anxiety. A similar explanation probably applies to the effects of anxiolytic drugs on behaviour in an unfamiliar open space. In a high stress version of the open field test, which includes loud white noise and bright light, anxiolytics increase rearing (McNaughton *et al.* 1984), whereas, in a low stress version of the test, anxiolytics decrease rearing (McNaughton 1985b). This result is explicable if rearing in this test reflects risk assessment in the same way that it does in the Blanchards' paradigms.

Note that the effects of the drugs in these cases cannot be attributed to an action on fear since they did not change the crucial behaviours in the Fear/Defence Test Battery. Nor can they be attributed to a direct effect on the motor systems which control risk assessment, or risk assessment would be uniformly decreased. Rather, they must be interpreted as a reduction in the underlying construct of anxiety. This is, by hypothesis, linearly related to defensive distance which is in turn, by observation, non-linearly related to risk assessment behaviour in undrugged animals. This non-linearity of the relationship between risk assessment and anxiety demonstrated by the drugs is a reason for preferring behavioural inhibition as the primary measure of this emotional state. With behavioural inhibition we see a simple progression from no threat, where there is no behavioural inhibition, to moderate threat, where there is behavioural inhibition of non-defensive but not defensive (risk assessment) behaviour, to high threat, where there is behavioural inhibition of both non-defensive and defensive behaviour.

2.3 THE ETHOLOGY AND ETHOPHARMACOLOGY OF ANXIETY AND FEAR

It is clear from the ethological analysis we have summarized that defensive responses to threat take many specific behavioural forms, each finely tuned to the precise circumstances of the threat. However, there are at least two global functional categories which the Blanchards (e.g. Blanchard and Blanchard 1990a,b, 1994) identify with fear and anxiety respectively.

Confrontation with an inescapable predator elicits an explosive attack (the precise details of which vary with the species and situation). This attack can be intermixed with undirected escape attempts (flight). A more distal/escapable predator also elicits escape

(now directed rather than undirected) or, if escape is not available, a complete absence of movement (freezing). A critical point, to which we shall return, is that all of these forms of behaviour can be seen as outputs from a system that controls behaviour when the animal is moving, or intends to move, *out of* a dangerous situation, i.e. when its goal is avoidance.

Where a predator is potential rather than actual (e.g. it has just gone out of view, or its smell is present), this elicits behavioural inhibition as the predominant response when the perceived risk is high. Behavioural inhibition (in the sense of inhibition of pre-predator behaviours) persists, but is accompanied by risk assessment behaviour when the perceived risk is moderate. Finally, risk assessment decreases when non-threat behaviour patterns are released from behavioural inhibition as the perceived risk becomes negligible. Risk assessment functions to reduce the level of perceived potential threat, while behavioural inhibition both prevents competition between risk assessment and non-defensive behaviour and reduces risk should threat prove actual.

Crucial to our theory is that both risk assessment and behavioural inhibition are part of a behavioural program which restrains the animal when it is moving, or intends to move, *into* a dangerous situation. They must operate only when the animal is faced with two incompatible goals (e.g. approach to food vs. avoidance of a predator). As we shall see, however, when we come to the full version of our theory, behavioural inhibition in the absence of observable external (behavioural) risk assessment can nonetheless be viewed as involving internal risk assessment (involving, for example, attentional shifts and the scanning of memory stores rather than the environment). Accompanying these changes in risk assessment are changes in heart rate and in the sensitivity to startling stimuli. Not only are these latter reactions sensitive to anxiolytic drugs in a way that predator-avoidance reactions are not, they also appear to be quite distinct ontogenetically (Hunt *et al.* 1994). As to the anxiolytic drugs, their central action is not on avoidance in the active sense of that term, but rather on the capacity of signals of danger to reduce approach to the appetitive goal.

2.4 SOME TERMINOLOGICAL PROBLEMS

We can now make a preliminary assessment of some of the problems involved in analysing anxiety. Anxiety can be viewed, to some extent as DSM-IIIR (American Psychiatric Association 1987) suggests, as a set of responses to threat. But theoretically, and ethologically, there are at least two quite separate types of threat. On the one hand, there is a definite, localizable, actual threat such as a predator, which must be avoided; and, on the other, there is indefinite, diffuse, potential threat, which (for other reasons, e.g. loss of food resources) must be approached. Both the functionally desirable and the observed skeletal responses are quite different depending on which of these two situations is present.

The Blanchards' analysis links actual and potential threat with 'fear' and 'anxiety' respectively. However, such a categorical distinction between fear and anxiety does not fit with the everyday tendency to conflate the two terms; nor with other ethologically based usage (e.g. Ursin 1980). There are also specific functional reasons for treating this

distinction with caution. First, neither the form nor the function of the behaviour is driven solely by the actual versus potential nature of the threat. For example, a potential threat which *need not be approached* requires only avoidance or escape, and so is functionally equivalent to a paradigmatic case of actual threat. Similarly, a localized actual *mild* threat which *has to be approached* (if some appetitive goal is to be obtained) will require behavioural inhibition and risk assessment, and so is functionally equivalent to a paradigmatic case of potential threat.[1] Second, and of particular relevance to our analysis of learning in the next chapter, even when there is an approach–avoidance conflict, escape responses are regularly observed. For example, in the Blanchards' paradigm case there are frequent rapid returns of the animal to its burrow. This escape from a potential predator is not only functionally equivalent to escape from an actual predator but is similarly insensitive to anxiolytic drugs (Appendix 1).

Thus, instead of focusing on the paradigmatic stimuli, use of which allowed the Blanchards to characterize the separate classes of behaviour and their functional consequences, we shall focus in the remainder of this book on the behaviour patterns themselves and their functions. Rather than refer (except in the loosest way) to fear, therefore, we shall refer to the fight–flight–freezing system (though usually dropping the 'freezing' term for the sake of brevity): a system presumed to control escape and active avoidance behaviour; to be homologous to the systems which control panic and simple phobia in human beings; and to be insensitive to anxiolytic drugs. Similarly, rather than refer (except in the loosest way) to anxiety, we shall refer to the behavioural inhibition system: a system presumed to control not only behavioural inhibition (in the sense of inhibiting prepotent behaviour) but also risk analysis; to be homologous to the systems which control anxiety in human beings; and to be sensitive to anxiolytic drugs.

Thus the behavioural inhibition system is a system which produces not only behavioural inhibition, but also risk assessment behaviour. In addition, in preparation for the sudden appearance of a predator, we would expect increased vigilance, faster reactions and a variety of more general autonomic changes (Fig. 1.1). One of these preparatory reactions may well be the basis for the phenomenon of 'fear-potentiated startle'. This consists of an enhanced startle response to an intense stimulus if this is presented in a threatening context (most commonly provided in animal experiments by a fear conditioned stimulus previously associated with a source of pain). Fear-potentiated startle has been much studied in the laboratory but is simple enough to have superficial ecological validity even there. In human beings potentiated startle has been shown to occur in the context of fear-provoking images, but not with affectively positive (sexual) images which produced similar levels of arousal (Jansen and Frijda 1994) or even with affectively negative (disgusting) images of a non-threatening type (Balaban and Taussig 1994; Kaviani *et al.*, 1999).

1. The threat must be only mild if approach is to be a viable option. The importance of this point will become clearer in the next chapter. But it is implicit in the Blanchards' emphasis on risk assessment that the reactions to the potential predator involve a major *conflict* between achieving the goal of food (and engaging in other survival-oriented behaviour outside the burrow), on the one hand, and achieving the goal of safety, on the other. Risk assessment behaviour can be seen, here, as active attempts to resolve this conflict by gathering sufficient information to resolve the approach–avoidance conflict either with avoidance (if a cat is detected) or approach (if a cat is proved to be absent).

This is not to say that we can always unambiguously distinguish the output of the fight–flight system from that of the behavioural inhibition system. Freezing occurs very clearly in response to a fairly close, inescapable predator, but something very like it can occur as the initial stage of the response pattern to a potential predator. The latter could be a high-intensity form of behavioural inhibition which is neurally distinct from freezing, but behaviourally difficult to distinguish from it because both involve cessation of movement. (Remember that 'behavioural inhibition', as we have defined it, is of prepotent responses. It does not imply loss of all behaviour and can be accompanied by active risk assessment.) It is unclear, therefore, whether stillness as such should be classified as purely a response to actual threat or also as a response to extreme potential threat. There are some indications of differential pharmacological sensitivity relating to the two (putative) types of response; and it may also be possible to distinguish them in operant paradigms (e.g. Kim *et al.* 1996) or via the specific body postures used to achieve stillness (Blanchard and Blanchard 1996). This issue is impossible to resolve on purely functional grounds and will, therefore, be set aside until a detailed analysis of neural mechanisms is presented (Chapter 6).

In this chapter, we have depended on the ethological study of unlearned defensive responses. This has given us a reasonable basis for the categorical distinctions we have tried to make in terms of the presumed functions of the behaviour concerned. A preliminary functional categorization of different emotional terms, derived from this analysis, is provided in Fig. 2.3. With the addition of ethoexperimental and ethopharmacological analysis, this approach has also given us some idea of the control of the relevant forms of behaviour. From this we have concluded that anxiety represents activity in the behavioural inhibition system: a system that is active when the animal is faced with a threat which it has some reason to approach, and the major outputs of which are inhibition of prepotent responses, risk assessment, and increased arousal or vigilance.

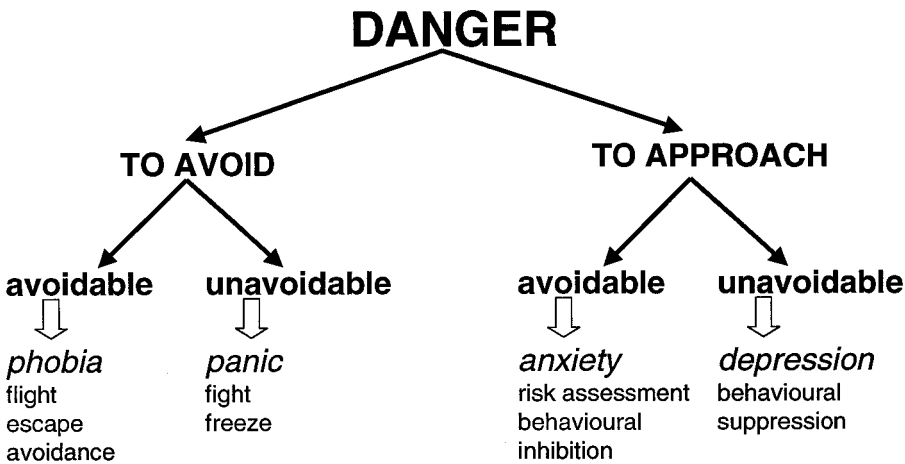


Fig. 2.3 A functional classification of different types of clinical disorder and normal response in terms of different classes of danger. Depression is not covered in this book. (Adapted from McNaughton 1993.)

The emphasis in this chapter has been on our evolutionary past. However, while many would accept that human emotion has evolutionary roots, most would argue that a key feature of both normal and pathological human fear and anxiety lies in the specificity of the eliciting stimuli. This specificity owes little to genetics and much to learning and cognition. We therefore turn to this issue in the next chapter.

3 Learning theory and anxiety

In the contemporary world, where cognitivism is running riot through psychology and neuroscience, it might seem somewhat old-fashioned to discuss learning theory—a child of extreme behaviourism. However, as Dickinson (1980) argues in his book *Contemporary animal learning theory*, 'the conditioning experiment [is] primarily an analytically tractable tool for studying the cognitive changes that take place during learning' (p. xi). Paradoxically, it is these same artificial and restricted conditioning experiments that have provided the clearest objective evidence that what changes during a rat's learning is not the strength of some internal link between a stimulus and a response, but rather knowledge of a goal (McNaughton 1989a, pp. 22–24). In this chapter, therefore, we will draw on the experimental rigour of the behaviour analyst to arrive at conclusions which will be useful in our analysis of the *cognitive* processes underlying anxiety.

The ethological analysis of Chapter 2 provided strong reasons to distinguish between a stimulus which warns of a present, localizable, proximal danger which must be avoided and a stimulus which warns of the possibility of (often non-localizable and distal) danger but which must be approached. This distinction was supported by the differential effects of anxiolytic drugs. The distinction between a primary aversive event and a signal which warns of that event is one which has also received, quite independently of ethology, considerable attention within learning theory. In particular, it is part of that body of data which supports a view of emotions as 'central' states (that is in the head, whether in the brain or in the mind) which are elicited by primary or secondary reinforcing stimuli (e.g. Mowrer 1960; Plutchik 1962; Millenson 1967). This view is usually embedded within a 'two-process' account of learning. The present chapter, therefore, discusses two-process theories of learning. Since we have dealt with these in considerable detail elsewhere (Gray 1972, 1975; McNaughton 1989a, Chapter 9), we provide only a brief summary below.

3.1 TWO-PROCESS THEORIES OF LEARNING

The two processes of two-process learning theory are those of classical ('Pavlovian', 'respondent', stimulus–stimulus, sign–significate, S–S) and instrumental ('Thorndikian', 'operant', stimulus–response, S–R) conditioning. Whether these processes are fundamentally distinct from each other has been controversial (e.g. Mackintosh 1974), although we believe there is good evidence that they are (Gray 1975; McNaughton 1989). Certainly, there is good reason to suppose that there are quite separate purely associative and reinforcement forms of behavioural plasticity, which could be supported by the neural phenomena of long-term potentiation (Bliss and Lømo 1973) and activity-dependent facilitation (e.g. Kandel and Hawkins 1992) respectively. In any particular situation, however, the behaviour actually acquired by the subject is normally the outcome of an

interaction between the two processes and may, in different species or experimental conditions, reflect one process more than the other despite the formal contingencies imposed by the experimenter.

Classical conditioning is the process whereby the subject learns the associative relationships between discrete stimulus events. Suppose a stimulus (S1) can be followed by a second stimulus (S2). Also suppose that S2 has a fixed probability (pS2) of occurring on occasions when it is *not* preceded by S1. There are now three possibilities for the conditional probability of S2 occurring given that it does follow S1 (pS2/S1): pS2/S1 may be greater than, equal to, or less than pS2. In classical conditioning experiments animals are sensitive to all three of these relationships (Dickinson 1980). Even in the case where pS2/S1 = pS2, and S1 apparently has no signalling value, this lack of signalling value is itself learned, producing 'learned irrelevance' (Mackintosh 1973).

In experiments using Pavlov's (1927) paradigm with pS2/S1 > pS2, it often seems that S1 (the conditioned stimulus, in this positive case conventionally designated CS+) comes to elicit a conditioned response (CR) by gaining control of responses which originally occurred to S2 (the unconditioned stimulus or UCS). However, it can be demonstrated that the CS–UCS association is formed independently of whether the UCS elicits any observable response (Dickinson 1980; but see Blackman 1983 for a contrary view), as would be expected from the properties of associativity underlying long-term potentiation (see, for example, Abraham 1988). It is also generally better (see below) to view the CR as being elicited by the anticipation of the UCS rather than necessarily consisting of any component of the UCR. This is also the case when pS2/S1 < pS2. Here S1 (designated CS–) may often seem to have little effect other than to reduce responses to a CS+; however, in a number of cases the animal produces a completely new response appropriate to the anticipated *omission* of the UCS.

Thus we can conclude that the animal learns the relationship (positive, neutral, or negative) between the stimuli and then, quite separately, produces a response which is generally adaptive given the anticipated situation. That is, the animal learns about a particular *goal* and may well use an innate or previously learned response to achieve that goal. It should be noted that the experimenter has no control, in such a straightforward Pavlovian conditioning experiment, over the nature of the CR, once the specific UCS has been chosen.

Instrumental conditioning is the process whereby the subject acquires new behaviour patterns which have the capacity to alter the frequency of its exposure to stimulus events; it can therefore be viewed as response–stimulus or R–S learning (although we shall qualify this view somewhat below). As with classical conditioning, there are three possible cases for the occurrence of S when it is preceded by the response, R. The experimenter can arrange that pS/R is greater than, equal to, or less than pS. However, since the subject rather than the experimenter controls the occurrence of R, we also need to consider the subsequent probability of occurrence of R, conditional on the R–S relationship: pR/(R:S). For each of the pS/(R:pS) cases we must consider whether the animal produces R so that pR/(R:S) is greater than, equal to, or less than pR (pR is usually referred to as the baseline response rate). The nine combinations which result from the 3×3 cases we have discussed (Table 3.1) can be simplified in a number of ways.

First, while the subject can learn that pS/R = pS (i.e. that the occurrence of the stimulus is completely independent of the occurrence of responding), this learning does not usually change pR/(R:S) or pR, and any effects of the conditioning procedure can probably be viewed as a special case of classically conditioned learned irrelevance. Where S is an aversive stimulus, as defined below, it has been claimed to produce a special type of learned irrelevance called 'learned helplessness', which Seligman (1975) has related to the emotional state of depression.

Second, if pR/(R:S) = pR (that is, there is no change in the probability of responding when S is made a consequence of responding) for either of pS/R > pS or pS/R < pS, it will be found that pR/(R:S) = pR also for the other (S would be what is usually termed a motivationally neutral stimulus). Taken together with our first simplification we have now dealt with the middle row and middle column of the 3 × 3 cases with which we started (Table 3.1).

Third, an important feature of the particular version of two-process theory which we prefer is that it takes the remaining 2 × 2 cases and treats their instrumentally conditioned effects as arising from changes in only two systems, one concerned with maximizing reward and the other with minimizing punishment.

Table 3.1 Instrumental reinforcing procedures with unconditioned reinforcing events (S). The items within the table and their abbreviations are defined by the intersection of the columns, which categorize the change in probability of a specific operant response (R) given the prevailing R–S contingencies (i.e. a relationship determined by the subject), and of the rows, which categorize the change in probability of S given the occurrence of R (i.e. a relationship determined by the experimenter). 'Rew' and 'Pun' indicate two specific classes of stimuli; + and – indicate the presentation or omission, respectively, of these stimuli consequent on a particular response

	pR/(R:S) > pR	pR/(R:S) = pR	pR/(R:S) < pR
pS/R > pS	Rew+ (approach)	Neutral	Pun+ (passive avoidance)
pS/R = pS	Learned irrelevance	Neutral	learned helplessness
pS/R < pS	Pun– (active avoidance)	Neutral	Rew– (extinction)

Consider the diagonals of Table 3.1. In the top left-hand corner we have the situation where a stimulus increases response probability when it is presented contingently upon that response. Empirically, it is usually the case that the same stimulus will also decrease response probability when its omission (or termination) is contingent on the response. In Skinner's (1938) terminology such a stimulus is a 'positive reinforcer', or more colloquially a reward. Thus the top left-hand corner and bottom right-hand corner of Table 3.1 may be logically independent but they are not independent in practice—the same physical stimulus can be used in the two cases. We indicate this similarity by using the same symbol 'Rew' for each and then adding + to indicate an increase in

probability of occurrence of the stimulus or – to indicate a decrease in its probability, consequent on the response.

Similarly, any stimulus which will decrease response probability when it is presented contingently upon that response (top right-hand corner of Table 3.1) will also usually increase response probability when its omission (or termination) is contingent on the response. Such a stimulus can be called a 'negative reinforcer' or punishment (abbreviated 'Pun' in Table 3.1). In general, any specific type of stimulus, within certain limits of intensity, duration, etc., will be regularly classifiable as Rew, neutral, or Pun.

We have (tacitly) supposed so far that, for a given animal species, there exist stimuli which may act without prior learning as Rew and others which may similarly act as Pun. However, these stimuli ('primary reinforcers') can also act as UCSs in a classical conditioning paradigm, as a result of which, in addition to any UCRs and CRs that ensue, they confer on CSs, which are not themselves initially rewarding or punishing, 'secondary' or 'conditioned' rewarding or punishing properties. It is a convenient fact for the simplifications we have been attempting that the reinforcing property of the secondary stimulus is the same (i.e. Rew or Pun) as that of the UCS with which it was paired as a CS.

The central states elicited by such secondary rewarding and punishing stimuli have played a key role in theories of emotion. Thus Mowrer (1960) treats 'hope' as the state elicited by secondary rewarding stimuli and 'fear' similarly as the state elicited by secondary punishing stimuli. It should be noted that these states which anticipate primary reward and primary punishment, respectively, form a necessary background for the bottom row of Table 3.1. For it is only when an animal anticipates reward (hopes) that the omission of reward ('frustrative non-reward'; Amsel 1962, 1992) can affect its behaviour; and only when it anticipates punishment (fears) that it can be affected by the omission of punishment ('relief').

For any particular imposed contingency between response and stimulus, Table 3.1 provides two different ways of parcelling out behavioural outcomes. One may focus on the specific direction of change in response (the columns) or on the specific class of stimulus (Rew or Pun), the increase (+) or decrease (–) in the probability of which gives rise to the changes in response. It is, of course, an empirical question as to which, if either, of these classifications reflects real distinctions in the way the nervous system processes information in different learning situations. However, the parsimony of Table 3.1 (and learning theories in general) depends on there being general mechanisms for increasing or decreasing response rate, respectively, which are each common to all reinforcers of that class. In the absence of specific contrary evidence, then, consistency requires us to parcel together Rew+ with Pun–, and Rew– with Pun+, on the basis of their effects on response probability, in just the same way and for just the same reasons that we parcel one Rew+ (e.g. food for a hungry rat) together with another qualitatively different Rew+ (e.g. water for a thirsty rat). It will be recalled that we placed the same emphasis on functional consequences as opposed to physical stimuli in the ethological analysis of the previous chapter.

Since we shall have much occasion to distinguish between Pun+ and Pun–, it is worth restating the distinction in a slightly different manner. 'Pun+' means that, if a specified response is made, then it will be punished; 'Pun–' that, if the response is made, impending punishment will be omitted. In still other words, successful avoidance of punishment

requires a passive avoidance response in the former case, but active avoidance in the latter. Similarly, 'Rew+' means that, if a response is made, then it will be rewarded; 'Rew–' that, if the response is made, it will not be followed by reward. (In the terminology of the first edition, Pun– and Rew– were represented as 'Pun' or 'Rew', each with a bar over the top.)

3.2 ELICITED REACTIONS IN LEARNING EXPERIMENTS

There are also empirical reasons for this particular form of parcellation, reasons which have been most extensively dealt with in two-process analyses of frustrative non-reward. In particular, the omission of an expected reward not only suppresses subsequent responding (so that Rew– is located in the right-hand column of Table 3.1), but it also elicits a range of behaviour patterns which overlap with those elicited by Pun+ (Wagner 1966; Gray 1967, 1975). Also, a stimulus paired with frustrative non-reward (CS-Rew–) becomes secondarily punishing in the same way as a stimulus paired with punishment (CS-Pun+).

This chapter (unlike the previous one) has until now ignored the eliciting properties of stimuli. The similarities between the eliciting properties of Pun+ and the eliciting properties of Rew–, especially since they are not complete, may not seem as surprising as they should. Pause for a moment and consider the hungry and sexually receptive female rat who will work to obtain both food and a male rat (Bermant 1961a,b, cited by Carlson 1980, p. 333). She clearly treats both a piece of food and a male rat as Rew+. However, the elicited responses supported by these different stimuli are quite different. The female rat will present herself for copulation to the male rat but will eat the food. She never attempts to eat her mate as he approaches her, nor adopt a sexually receptive posture to the food. Given the different eliciting properties of these stimuli which have in common the reinforcing property of being Rew+, the fact that the eliciting properties of Pun+ and Rew– are similar is surprising.

The similarity between the eliciting properties of Pun+ and Rew– is even more remarkable when we consider that this is not a general property of negatively reinforcing stimuli. Primary and secondary aversive stimuli (Pun+ and CS-Pun+), respectively, elicit responses through quite different systems. For example, in the rat, an electric shock elicits as a UCR a great deal of activity (running, jumping, etc.) and noise (squealing); but a CS which predicts such a UCS actually elicits the reverse—immobility (freezing) and silence (Myers 1971). Similarly, a shock may elicit aggressive behaviour if a suitable object of aggression is present, but a CS paired with shock inhibits aggressive behaviour (Baenninger 1967; Myers 1971).

In this respect, aversive stimuli provide a clear contradiction to the idea that properties acquired by a CS as a result of classical conditioning are properties also possessed by the UCS employed to establish the CR. Thus aversive stimuli do not conform to Pavlov's (1927) stimulus substitution hypothesis. This is a particularly powerful demonstration of the fact that the CS alerts the animal to a particular goal rather than simply eliciting the UCR. These results should surprise only the hard-line S–R theorist, as opposed to the cognitive learning theorist (e.g. Dickinson 1980). They might, in contrast,

be predicted, and to some degree in specific behavioural detail, by the ethologist. For, as we saw in Chapter 2, the functional requirements for responses elicited by a predator or a warning of a predator change markedly depending on whether the source of danger can simply be avoided or must be approached. These distinctions were made on the basis of ecologically valid reinforcing stimuli. Nonetheless, on the reasonable assumption that a shock can be equated with contact with a predator, then a CS-shock should be equated with a potential or close predator, depending on whether it must be approached or avoided, respectively. In the former case, following the logic developed in the previous chapter, we expect anxiety to be elicited.

The key conclusions from the above considerations, then, are that: (a) we can to some extent equate Pun+ with Rew–, as well as CS-Pun+ with CS-Rew–; but that (b) we need to distinguish carefully between the primary events (Pun+, Rew–), on the one hand, and CSs for those primary events, on the other, since they can have quite opposite eliciting properties and functional requirements.

3.3 THE BEHAVIOURAL INHIBITION SYSTEM REVISITED

We are now in possession of the main distinctions which will allow the behavioural inhibition system, introduced in the previous two chapters, to provide the core of our theory.

As originally proposed by Gray (1975), the behavioural inhibition system mediates certain responses to secondary punishing and secondary frustrative stimuli (CS-Pun+, CS-Rew–; Table 3.1). The similarities between the responses to these two classes of stimuli (Wagner 1966, 1969), discussed above, are then attributed to their acting on the same central system. This system—the behavioural inhibition system—governs approach to, coupled with passive avoidance of, stimuli which warn of danger, either innately (e.g. cat odour) or as a result of the conditioning process that gives rise to CS-Pun+ or CS-Rew–. Primary punishing and primary frustrative stimuli (Pun+, Rew–) are similarly held to act on the same system as each other, the fight–flight system (Gray 1971, 1975) or, in full, the fight–flight–freeze system of Chapter 2. The fight–flight system, as we have seen, is quite distinct from, and to some extent has opposite properties to, the behavioural inhibition system. These two systems have in common, however, the property that neither is activated by primary reward stimuli (each subclass of which acts on its own specific appetitive system) or by secondary reward stimuli and safety signals (CS-Rew+; CS-Pun–); these all act on a third system, the behavioural approach (Gray 1972; Gray *et al.* 1991) or activation (Fowles 1980) system.

As noted above, the distinction we have made here between Pun+ and CS-Pun+ can be mapped directly onto the one made in Chapter 2 between actual and potential threat. Here, we take this argument one step further to claim that the parallel between these two distinctions is not mere analogy, but reflects an identity. This is justified on both functional and ethological grounds.

From the functional point of view, a CS-Pun+ is a potential threat stimulus, since it involves an explicit conflict between a goal which would normally elicit a prepotent approach response and stimuli which warn of danger should that goal be approached. (If there is no prepotent approach tendency then there will be no occasion when the

Pun+ contingency can be demonstrated.) This is true despite the fact that the animal depends on conditioning rather than evolution for both the nature of the prepotent response and for the nature of the warning stimulus. From the ethological point of view, a CS-Pun+ normally releases the same behaviour patterns as does an innate potential threat stimulus such as cat odour. To complete the terminology developed in the present chapter, we can therefore code such an innate potential threat stimulus as 'IS-Pun+' (IS standing for 'innate stimulus').

It follows from the above argument that the behavioural inhibition system should react to both conditioned aversive stimuli (CS-Pun+, CS-Rew–) and innate ones (IS-Pun+, IS-Rew–). For human beings, darkness could be one IS-Pun+; in our phylogenetic past it often contained (and in many modern cityscapes still does) a plenitude of potential dangers. Whether there are any true IS-Rew– is not clear. Care must in all cases be taken in the classification of innate stimuli, unless their functional status is very obvious. In a conditioning paradigm, we have experimental control over the functional status of any particular stimulus and so can usually determine whether it is Pun+ or CS-Pun+. In contrast, with innate stimuli, all we can determine from conditioning experiments is that they are aversive. It is clear enough, however, that a bite from a predator can be equated with Pun+. A close approach from the predator (involving no pain but requiring avoidance) should probably, then, be treated as functionally equivalent to a CS-Pun–, i.e. it is an 'innate fear stimulus' (Gray 1971, Chapter 2; Gray 1976), for example, snakes for primates (Hebb 1946). The odour of the predator (involving no pain, warning of danger, but often occurring in a context that requires approach) should in many cases, therefore, be equated with CS-Pun+. But careful functional analysis (as in the Blanchards' work on cat odour discussed in Chapter 2) is required to be sure that any particular stimulus is an innate anxiety stimulus (IS-Pun+) rather than an innate punishing stimulus (Pun+). To repeat: the key issue in all cases will not be simply whether the stimulus signals actual or potential danger but rather the extent to which other aspects of the situation require approach or avoidance.

In addition to innate signals of potential threat such as cat odour, which will retain their warning value even after many presentations, the same functional considerations lead to the view that the behavioural inhibition system should react to at least certain forms of novelty. On first encounter, stimuli such as unexplored environments should be treated as sources of potential danger and should be dealt with in the same manner as a CS-Pun+, CS-Rew–, or IS-Pun+. There are two critical distinctions between such novel stimuli and true innate anxiety stimuli. The first is that after a period of habituation they become neutral stimuli. Thus the phenomena described by Sokolov (1963) as the 'orienting reflex' and by Pavlov (1927) under the rubric of 'external inhibition' should be closely related to the activities of the behavioural inhibition system. The second is that novelty has the approach and avoidance components of conflict both instantiated in the same stimulus. A novel object or environment should be treated not only as a potential source of danger, but simultaneously as a potential source of reward. Unlike other aversive stimuli, therefore, it does not require any concurrent appetitive stimulus to generate the required conflict.

At this point in the development of the model, then, we view the behavioural inhibition system as being activated in any approach–avoidance conflict; we equate anxiety

with activity in the behavioural inhibition system; and we view anxiolytic drugs as having a selective effect on the behavioural inhibition system. In Chapter 10, the full theory will slightly expand on the specification of adequate inputs to the behavioural inhibition system. In the interim, until we have considered the pharmacology and neurology of anxiety, we can take anxiety-provoking stimuli to be: CS-Pun+ ('conditioned fear stimuli'), CS-Rew– ('conditioned frustrative stimuli'), IS-Pun+ such as cat odour ('innate anxiety stimuli'), and some but not necessarily all novel stimuli. With all of these stimuli, their definition implies that they activate the behavioural inhibition system, not because they are aversive as such, but because the animal has some additional tendency to approach rather than simply avoid them. *Given an innate or learned requirement to approach the source of potential danger*, the animal would normally be advised to inhibit non-defensive behaviours (including the prepotent approach response) as well as the fight–flight system (provided this is not extremely highly activated). Then, while maintaining this behavioural inhibition, it should actively assess the degree of threat while at all times actively scanning/assessing its environment. All of these responses also imply increased arousal and attention in the sense of heightened readiness to convert approach (requiring activity in the behavioural inhibition system) to avoidance (mediated by the fight–flight system) as fast as possible should conditions require this.

This analysis suggests the system of Fig. 3.1. There are a number of key points to be borne in mind when interpreting this figure.

First, while we have included CS-Pun+ within the class of anxiety provoking stimuli, we have explicitly excluded Pun+. For some, it is CS-Pun+ which elicits 'fear' and we have explicitly recognized this by referring to these stimuli as 'conditioned fear stimuli' in Fig. 3.1. However, for others such as the Blanchards (Chapter 2), fear is a state elicited by Pun+. We can acknowledge this usage by referring to these stimuli as 'unconditioned fear stimuli'. As noted above, however, conditioned and unconditioned fear stimuli usually

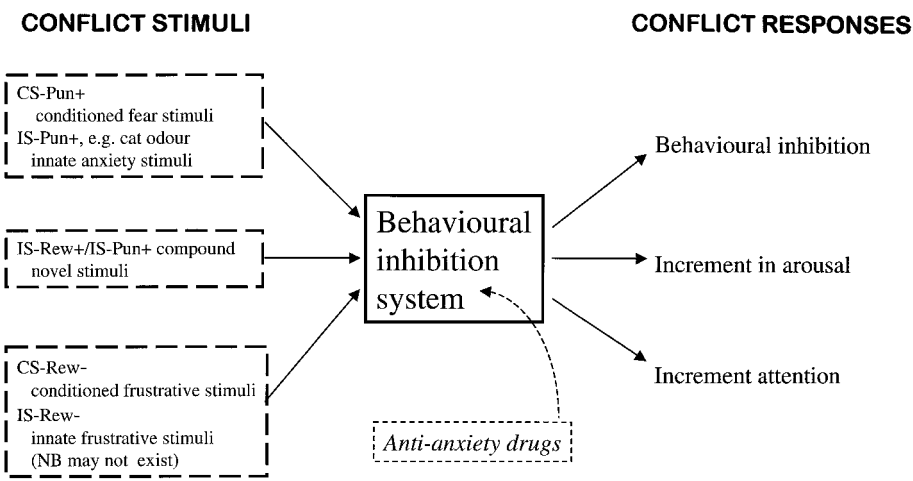


Fig. 3.1 The behavioural inhibition system—expanded. Re-specification of the inputs of Fig. 1.1 in terms of the learning theoretic analysis of the current chapter.

have quite different eliciting properties, resulting in behavioural inhibition and escape respectively; and, as we shall see later, they have quite different pharmacologies. It follows that conditioned and unconditioned fear are often significantly different emotions, and it may therefore seem problematic to label them both 'fear'. We shall attempt to avoid using the term as far as possible (replacing it with, for example, activity in the fight–flight system or behavioural inhibition system as appropriate); and, when we are forced to use it, we shall preface it with 'conditioned' or 'unconditioned'. However, conceptually and at the level of neural analysis, there is no real problem. By our current definition, anxiety involves an approach–avoidance conflict. This necessarily implies that fear systems could be activated; but pure fear can be contrasted with anxiety because, in the latter case, the tendency to avoid (based on fear) is combined with a tendency to approach. As we shall see, this leads naturally to the idea of levels of control of the defence system; anxiety then reflects activity in a higher order system which modulates the output of a more fundamental pure fear system.

Second, while we have included CS-Pun+ (i.e. stimuli which generate an approach–avoidance conflict) within the class of anxiety-provoking stimuli, we have explicitly excluded CS-Pun– (i.e. safety signals). A stimulus which predicts that a particular response will escape or avoid an aversive stimulus would be presumed to activate the behavioural approach system and therefore to generate conditioned relief. However, precisely the same stimulus can, and often will, predict that some other (usually many other) responses will produce, or fail to avoid, the aversive stimulus. In many situations, therefore, particularly during the initial phases of active avoidance conditioning, conditioned fear (elicited by CSs paired with the primary aversive UCS) is likely to participate in the processes leading to both anxiety (engendered by CS-Pun+) and relief (dependent on CS-Pun–). That is, both the behavioural inhibition and the behavioural approach system will be activated concurrently, with some patterns of behaviour being produced by the one system and some by the other.

Third, at the stimulus level, we have mapped a distinction derived from learning theory (the difference between, on the one hand, Pun+ and Rew–, and, on the other, CS-Pun+ and CS-Rew–) onto a distinction derived from ethology (the difference between actual and potential threat). However, the reverse mapping is not possible. The experimental difference between Pun+ and CS-Pun+ is simply one of conditioning. However, the ethological point of view contradicts the idea that anxiety can be defined uniquely in terms of this conditioning. In particular, the odour of a cat, as used in the experiments by the Blanchards discussed in the previous chapter, is an innate anxiety stimulus. Thus the commonly perceived relationship between fear and innate stimuli, on the one hand, and anxiety and conditioned stimuli on the other (a relationship much emphasized in the first edition of this book), is an artefact of conventional laboratory methodology rather than a logical necessity.

Fourth, this same ethological analysis resolves an apparent problem with the categorization of stimuli provided by the first edition of this book. There, anxiety-provoking stimuli were defined as the aggregation of: signals of punishment, signals of non-reward, novel stimuli, and innate fear (the term then used) stimuli. In this set, if the definition of anxiety and the behavioural inhibition system requires the involvement of conditioning, both novel and innate fear stimuli constitute *ad hoc* additions to signals of

punishment and reward omission. However, given that we have already distinguished ethologically between threat avoidance (depending upon the fight–flight system) and threat approach (producing conflict and hence risk analysis and behavioural inhibition, mediated by the behavioural inhibition system), there is no problem in identifying CS-Pun+ and CS-Rew– as anxiety-provoking stimuli and adding them to the novelty and innate anxiety stimuli already identified in the ethological analysis.

From an ethological point of view, CS-Rew– is not so easy to classify, although such stimuli can be readily identified within the experimental tradition of learning theory. The strongest reasons for including them with the other anxiety-eliciting stimuli lie in the similarity of their reinforcing and eliciting properties to those of CS-Pun+. However, *post hoc* it can be suggested that, under normal ecological circumstances, loss of food can be just as life-threatening as predation, and that anticipated loss of food requires behavioural inhibition and risk analysis in the same way as potential predation.

Fifth, while the ethological analysis corrected some inappropriate inferences from the learning theory analysis, the latter in turn highlights an important point about the coactivation of anxiety and relief mechanisms in the ethological situation. In the situation studied by the Blanchards (Chapter 2), after prior experience of a cat, rats inhibit their exit from the burrow system and, when they exit, engage in scanning, this behaviour being motivated by anxiety. However, when the rat has entered the arena and then chooses to escape back into the burrow system, this is escape from the same source of fear, but now motivated by anticipatory relief or hope (i.e. by the response to the equivalent of a CS-Pun–, namely, the burrow). Thus (as we have already argued for the case of avoidance conditioning), in the arena responses will be activated (or primed) which are appropriate to both an actual and a potential predator, and the animal can rapidly switch between these repertoires by releasing one or the other system (see McNaughton 1989a, Chapter 2).

Sixth, not all stimuli which technically fit the specific labels of Fig. 3.1 will always be anxiety provoking. A novel stimulus which the animal has no innate basis for perceiving as potentially threatening (e.g. a smell of roses) will be different from one for which there may be such a basis (e.g. a loud noise). Similarly, what is formally a CS-Pun+ from the experimenter's point of view will be anxiety provoking only if the rat truly anticipates punishment, as would, for example, be the case in a conditioned suppression paradigm where the shock is unavoidable. However, in a well-learned avoidance task, habitual suppression of responding in the presence of the CS means that the shock is seldom or never obtained. In this case the CS should no longer produce anxiety.

As was the case with the ethological distinctions made in Chapter 2, experiments with anxiolytic drugs (McNaughton 1985c) support the distinction just made in relation to conditioning paradigms. Anxiolytic drugs given before every training session impair the acquisition of both successive negative discrimination (a paradigm in which behavioural inhibition to a CS-Pun+ results in shock avoidance) and conditioned suppression (which is similar except for the fact that shock cannot be avoided). However, if the drug is administered after suppression has been acquired and the dose is increased progressively rather than suddenly, then anxiolytics impair performance of conditioned suppression but not successive negative discrimination. This supports the inference that, provided a CS-Pun+ warns that a shock is likely to occur, it will generate anxiety; but that once

behavioural suppression is well developed and shocks no longer likely, then such a stimulus no longer generates anxiety. Similar conclusions have been drawn for changes in fear after prolonged training with CS-Pun– (Kamin *et al.* 1963; see McNaughton 1989a, p. 125). Indeed, it seems generally to be the case that, once a consistent pattern of active or passive responding to an aversive or appetitive reinforcer is established, not only do the eliciting properties of the CS disappear, but the response itself may become a simple habit unrelated to the value of the original reinforcer even if, as in the case of positive reinforcers, the reinforcer continues to be delivered (see McNaughton 1989a, pp. 124–128).

3.4 CONCLUSIONS

We believe that Blanchard and Blanchard (1990a) are right that functional analysis is required if we are to understand the unity of the various different forms of behaviour labelled 'anxious'. However, the methodologies of the two-process theorists, with which the Blanchards contrast their 'ethoexperimental' methodology, provide an independent basis for categorizing and, in particular, generalizing. While ethology may tend to overcategorize, learning theory tends to overgeneralize. The combination of the two provides a better balanced and potentially deeper perspective. We apply this perspective to the symptoms and syndromes of anxiety and anxiety-related disorders in Chapter 11.

Whether we consider the learning theory analysis of the present chapter or the ethological analysis of the previous one, the anxiety we have defined has proved specifically sensitive to anxiolytic drugs. Indeed, in a number of cases, we have seen that the effects, or lack of effects, of anxiolytic drugs in specific tests have clarified their psychological analysis. In particular, we have found that variations in apparently trivial parametric aspects of experimental procedure (e.g. the extent of training) which do not change the superficial behavioural results, nonetheless often dramatically change the underlying psychological processes involved. In what follows then, we shall use the anxiolytic drugs as a probe, or marker, for activity in the behavioural inhibition system and hence for its neural location. Our next step, therefore, will be to review the behavioural effects of these drugs. These can then be related to the behavioural effects which would be expected from damage to specific parts of the defence systems of the brain.

4 The anxiolytic drugs

Where should one should draw the line between primary defence systems and systems mainly involved in anxiety? Our principal tool for making this distinction will be the results of a careful comparison (Appendix 1) of the profile of behavioural effects of the anxiolytic drugs with that of specific lesions of each potentially relevant structure. This strategy relies upon the assumption that anxiolytic drugs are likely to act in the same direction as a lesion of those neural structures most involved in anxiety. As a necessary preliminary to its execution, this chapter surveys the drugs concerned and their clinical and behavioural effects.

4.1 WHAT ARE THE ANXIOLYTIC DRUGS?

In the previous two chapters we occasionally used the selective actions of the anxiolytic drugs as an additional, and quite different kind of, justification for distinctions already made on ethological or learning theory grounds. The fact that any class of drugs consistently affects behaviour patterns from one ethologically or learning-theory defined class while having no effect on behaviour from the other class provided additional evidence for the correctness of that classification. The fact, moreover, that the class of drugs concerned was defined solely by their clinical effectiveness in treating anxiety disorders added weight to the ethological and psychological characterization of the behaviours as reflecting anxiety. Furthermore, the consistent effects of the drugs provided us with additional reason to link the separate ethological and learning-theory classifications. However, any type of drug which produced the relevant dissociations would have been satisfactory for this purpose. The fact that it was the quite independently defined anxiety-related class of behaviour that is affected by the anxiolytic drugs provided evidence, in return, that the clinical definition of the drugs as 'anxiolytic' was correct. Further, we found that the effects of the drugs were best characterized as a reduction in the underlying construct of anxiety rather than a simple reduction in specific anxiety-related behaviours, such as rearing or response suppression.

In this chapter we take it as given that anxiolytic drugs do in fact reduce anxiety as assessed in the clinic. We then go on to clarify anxiolytic drug action as comprising that subset of the effects of each clinically effective drug which are common to all classes of clinically effective anxiolytics. We shall also provide more evidence that these drugs act on the behavioural inhibition system defined in the two previous chapters. This is an important precursor to our use, in the remainder of the book, of the anxiolytic drugs as a 'marker' (rather like the use of a radioactive label in biochemistry) for psychological processes which contribute to anxiety and for the neural systems subserving those processes.

There is now a vast literature on anxiolytic drugs (see Nutt 1990, for a review of data on human subjects). Here we summarize the conclusions justified in detail in

Appendix 1. Even there, we depend heavily on secondary sources, particularly Gray (1977) for the classical anxiolytics, and Griebel (1995) and Handley (1995) for compounds acting selectively at the 5-hydroxytryptamine (5HT) 1A receptor and imipramine. We shall not include compounds that are presumed from some animal test or another to be anxiolytic, but which have not been tested with human patients. We first look briefly at the clinical use of anxiolytic drugs in human beings to define our target compounds, and then consider in depth the effects of those compounds in animals. An important feature of this two-part approach is that we shall first use the clinical data to provide us with a preliminary indication of which drugs are anxiolytic (a definition assumed until now). Then, given this ostensive definition, we shall use the animal data as a basis for discovering the psychopharmacology, and in later chapters the neuropharmacology, common to these drugs, treated as a single functional class.

4.2 CLINICAL USAGE

Using the systemic administration of any drug as a tool for the dissection of specific psychological processes is subject to a major problem: all drugs have a wide range of disparate effects. It can be difficult, therefore, to tell whether a specific behavioural action of the drug is due to its main effect, in which we are interested, or to some side-effect. In this respect the present edition is on a much sounder footing than the first edition. The recent development of 'novel' anxiolytic drugs, which share none of the clinical side-effects of the classical anxiolytics, allows us to compare the behavioural effects of the two classes. Where they overlap we can, as we shall see, be almost certain that the effects reflect their anxiolytic action. Where they differ, we can ignore the separate effects as being due to side-effects.

So, what are the anxiolytic drugs? We have already had a brief look at the definition of anxiety and its subtypes (Chapter 1). In the ideal case, an anxiolytic drug should act to reduce anxiety as defined by, for example, DSM-IIIR (American Psychiatric Association 1987), while affecting no other psychological state or syndrome. As we shall see, when these compounds are taken as a group but not drug by drug, this is largely the case. Of particular importance in this context is the fact that, despite considerable debate as to which particular drug is best for the treatment of anxiety in general (see, for example, Goldberg 1990; J. Marks 1990; and their accompanying commentaries on each other), no one of the many different anxiolytic drugs is usually seen as being completely specific to one or more of the DSM-IIIR subtypes of anxiety. The issue is complicated to some extent by drugs which, in addition to having a general anxiolytic action, are particularly good at reducing panic attacks or obsessions. However, as argued in Chapter 1 (and again in more detail in Chapter 11), panic and obsessions proper are likely to be neurally quite distinct from the anxiety which can accompany them. We shall, therefore, discuss anxiolytic drugs on the working assumptions that they all treat all types of anxiety, and do so equally well given proper adjustment of dose.

There is one point at which we depart from the DSM-IIIR classification: phobia. We have already seen that, on ethological grounds, phobia should not be seen as part of anxiety, and that this separation is consistent with the failure of anxiolytic drugs to

prevent avoidance of a phobic object such as a snake (see Sartory *et al.* 1990, p. 273, for a brief review). Further, phobia does not activate the areas of the brain which we shall associate with anxiety, nor are changes in the pattern of brain activity produced by a phobic stimulus altered by anxiolytic drugs (Fredrikson *et al.* 1995).

It is important for the approach of this book to show not only that anxiolytics can treat anxiety, but also that they are relatively specific in such treatment. If anti-anxiety drugs also reduced joy, anger, and the sex drive, we should make little progress in deducing anything about anxiety from studying them. Similarly, if the benzodiazepines, for example, were as effective in controlling schizophrenic thought disorder or manic elation as in alleviating anxiety, this would make them therapeutically very useful, but greatly limit their value as scientific tools. Fortunately for the scientist, the relative specificity of anxiolytic drugs as a class is clear. They are of little use in schizophrenia or mania. Nor are they effective in the control of anger or aggression; indeed, they can facilitate aggression (Lynch *et al.* 1975).

Of course, each particular chemical type of anxiolytic has its own non-specificities, which are dealt with below. But, as a class, the anxiolytic drugs are specific. This specificity is least clear in relation to depression. It is in any case clinically difficult to separate anxiety and depression, which more often than not coexist. Some anxiolytic drugs can be effective in some types of depression (Frith *et al.* 1979; Kellner *et al.* 1979; Johnstone *et al.* 1980); and conversely, some drugs which are principally thought of as antidepressants are effective also in the treatment of anxiety (Frith *et al.* 1979). To leap ahead somewhat, however, we shall conclude that some, but critically not all, anxiolytic drugs have additional, functionally independent, actions on some types of depression.

So, what are the actions of the anxiolytics in the clinic? It has been said that all drugs go through an essentially similar clinical life cycle. On introduction to medical practice they are hailed as wonder-drugs, able to cure 100 per cent of cases of the target disorder (as well as, sometimes, everything else). Next, it is reported that they are totally without effect on the original target disorder and, far from curing other disorders, produce a range of life-threatening side-effects. Finally, it becomes accepted that in a subset of cases of the target disorder the drug can be helpful, at least in controlling symptoms, but doses must be controlled with care and precautions taken. The drugs generally classified as anxiolytic have followed this pattern, inasmuch as the classical anxiolytics have fallen out of favour because of their side-effects. But it is interesting to note that they have not been replaced by drugs with any greater or more general therapeutic main effect; indeed, somewhat the reverse appears to be the case for the novel anxiolytics. As a result, pharmaceutical companies have been driven to produce many chemically unrelated compounds, all of which share the same essential main effects.

There have been five major changes in drug type in the clinical treatment of anxiety in general and generalized anxiety disorder in particular. The oldest anti-anxiety drug of all is alcohol. Clinically this is unattractive because of its depressant, toxic, addictive, and intoxicant effects. Second, came the barbiturate family. These were the earliest drugs used regularly for the specific treatment of anxiety. Third, for a brief period came 'Miltown' (meprobamate). However, each of these has, albeit to a lesser extent, the same problems as alcohol. Fourth, sweeping the others into the dustbin of history, came the benzodiazepines, with much less pronounced sedative effects, a wide margin of safety

between the clinically effective dose and the dose at which you can kill yourself, and, it was wrongly thought, little potential for producing dependence and withdrawal symptoms. Fifth, the benzodiazepines themselves are now being challenged by novel classes of compound such as the azapirodecanediones, buspirone and ipsapirone, and by novel uses of older drugs, particularly the tricyclic antidepressant, imipramine.

As we shall see, there is good reason to group buspirone and imipramine together in terms of their agonist actions on $5HT_{1A}$ receptor systems. There is also a range of very new drugs which appear to be variants on either the benzodiazepine or the $5HT_{1A}$ theme. In both cases the newer compounds represent an attempt to retain the main effect of the parent compound while eliminating unpleasant side-effects. Thus the newer, apparently more specifically anxiolytic, benzodiazepines are partial agonists (this seems more important than the type of benzodiazepine receptor to which they bind) which retain the anxiolytic effect of benzodiazepine receptor activation, but essentially act as their own antagonists with respect to the depressant and addictive effects (Haefely 1990a,b, 1991; Dobel and Martin 1992; Facklam *et al.* 1992; Sanger *et al.* 1994). Similarly, ipsapirone is like buspirone in being effective in treating generalized anxiety disorder while lacking the effect of buspirone on dopaminergic systems (Cutler *et al.* 1993). Since these more recent, and apparently more specific, drugs all appear to share the actions common to the classical (e.g. benzodiazepines) and novel (e.g. buspirone) anxiolytics, we have largely excluded them from discussion. There are also at present insufficient data to draw any clearly different conclusions from those we draw for the classical and novel anxiolytics as groups. Finally, we shall briefly discuss below the use of drugs which block β-adrenergic receptors ('beta-blockers') in the control of anxiety; but we conclude that their action is primarily peripheral and so do not group them with the other anxiolytics in our search for the neuropsychology of anxiety.

There are indications that the neuropeptide cholecystokinin (CCK) may be panicogenic (Harro *et al.* 1993); that CCK antagonists may have anti-panic activity (see Crawley and Corwin 1994); and that the effects of certain benzodiazepines may be through direct or indirect interaction with CCK receptors rather than through benzodiazepine receptors (Boden and Woodruff 1994). But there are no data at present on clinical trials of CCK antagonists on either panic disorder or, more importantly for our present analysis, generalized anxiety. Animal behaviour tests suggest that neuropeptide Y may have some anxiolytic activity, but the critical receptor (Y-1) appears to be absent from the human brain (see Gehlert 1994). For both these peptides we need a much greater volume of data, and particularly clinical trials on generalized anxiety, before they can be included in the present type of analysis.

There are indications that $5HT_2$ antagonists may be anxiolytic (e.g. Hensman *et al.* 1991; Katz *et al.* 1993). Their effects appear to be like those of buspirone. How specific the drugs are, even within the 5HT receptor classification, is not clear; and, unfortunately, they have not been submitted to a wide range of behavioural tests. We have therefore excluded them from consideration.

Finally, there are suggestions that ondansetron, a $5HT_3$ antagonist, is effective in generalized anxiety disorder; and that, if so, it may be the first member of an entirely new class of anxiolytic drug which could be particularly free of side-effects (e.g. O'Hanlon

et al. 1995). However, no fully documented report of a clinical trial of the effects of ondansetron on generalized anxiety disorder has yet been published. Furthermore, since ondansetron is marketed primarily as an anti-emetic in cancer therapy, it is odd, given the anxiety that such patients must be feeling, that there appears to be no incidental report in that literature of anxiolytic action of the drug. Indeed, there is one report of a patient who experienced an apparently drug-related 'anxiety attack' (Sledge *et al.* 1991). This latter result could, however, be due to $5HT_3$ antagonists having a $\cup$-shaped dose–response curve (see Barnes *et al.* 1992, p. 108). A randomized double-blind placebo-controlled study of the related $5HT_3$ antagonist tropisetron (ICS 205-930) reported positive (but not dose-related) effects on generalized anxiety disorder (Lecrubier *et al.* 1993). However, until the pharmacology of tropisetron is more extensively known it will be difficult to evaluate this result, since the anxiolytic effects of nominal $5HT_3$ antagonists may well be due to action at some other receptor (see Barrett and Vanover 1993, p. 6). In sum, 'an anxiolytic effect of $5HT_3$ antagonists still needs to be demonstrated in clinical practice; however, a wide spectrum of activity does not seem likely. A robust anxiolytic-like effect has not been confirmed in any of the commonly used [animal] models for anxiety' (Broekkamp *et al.* 1989, p. 6). Until further data on their pharmacology and clinical effectiveness in generalized anxiety disorder are available, therefore, we have chosen not to treat $5HT_3$ antagonists as anxiolytic drugs, although we mention their effects occasionally below.

4.3 THE CLASSICAL ANXIOLYTICS

Ethanol, barbiturates, meprobamate, and benzodiazepines can be grouped together as the classical anxiolytics, which we shall contrast as a group with the more recently introduced novel anxiolytics. But, before we do so, it is important to note some of the similarities and differences within the group of classical anxiolytics themselves. Chemically they are very different from each other and produce their effects not through any common direct receptor action but through a final common pharmacological path. The benzodiazepines are marginally more effective than the barbiturates at treating anxiety (Rickels 1978) and are certainly much less toxic, but in virtually all other respects their similarities to the barbiturates far outweigh their differences.

The benzodiazepines include a number of drugs in exceptionally wide use, such as Librium (chlordiazepoxide) and Valium (diazepam). More than sixty million benzodiazepine tablets were prescribed annually in the United States alone in the 1970s (Tallman *et al.* 1980). This was

> the decade of peak usage, [but] alarm was expressed about the 'relentless march of the psychotropic drug juggernaut' and the benzodiazepines were stigmatized . . . as 'the opium of the masses'. . . . In the [1980s], 'a dark shadow has been cast over the benzodiazepines by the spectre of dependence' (Lader 1990, p. 3).

It is now generally agreed that benzodiazepines should be restricted to only short-term treatment and as an adjunct to other therapies (see review by Ashton 1994).

The classical anxiolytics as a group all share the property of enhancing the effects of the inhibitory neurotransmitter, GABA. They do so through quite different primary actions, each of which modifies the functioning of the 'GABA–benzodiazepine-chloride ionophore' complex through a slightly different route (Fig. 4.1B). The benzodiazepines bind to the benzodiazepine receptor, which is linked to the $GABA_A$ receptor, and so increase the affinity of the latter for GABA (Guidotti *et al.* 1978; Fig. 4.1A). For general reviews of benzodiazepine receptors and their ligands, see Haefely (1990a,b,c; Haefely *et al.* 1990; Richards *et al.* 1991). The barbiturates also increase GABAergic transmission (Ransom and Barker 1976; Barker and Ransom 1978), but by binding to the picrotoxin receptor, which is closely linked to the chloride channel, hence prolonging its opening (Olsen *et al.* 1978a,b; Ticku and Olsen 1978; Braestrup and Nielsen 1980). Ethanol appears to interact with yet another receptor linked to GABAergic transmission (Suzdak *et al.* 1986; see also Coop *et al.* 1990), as well as with receptors for the excitatory amino acid transmitter *N*-methyl-D-aspartate (NMDA). Interaction with the latter does not, however, appear to be the basis for the anxiolytic effects of ethanol (Criswell *et al.* 1994). Ethanol's interaction with GABAergic transmission may vary from brain region to brain region (Soldo *et al.* 1994). The lack of toxicity of the benzodiazepines, relative to the other classical anxiolytics, can probably be attributed to the fact that they are the only group which modulates the interaction of GABA with its receptor as opposed to interacting directly with the chloride channel (Fig. 4.1). They only have an effect, therefore, when GABA is released endogenously.

These properties of the classical anxiolytics give rise in a straightforward manner to the hypothesis that the necessary common denominator of anxiolytic action lies in enhanced GABAergic transmission. However, such a 'GABA' hypothesis of anxiolytic action suffers from the problem that GABAergic neurons and receptors are distributed far too widely in the brain for it to be plausible that their chief function is to regulate anxiety. Even if we restrict ourselves to those GABA receptors which also have benzodiazepine receptors linked to them, we find them in many regions of the central nervous system that are unlikely to play a role in anxiety. Thus, for example, the cerebellum and spinal cord are both rich in benzodiazepine receptors (Braestrup and Squires 1977; Möhler and Okada 1977; Robertson *et al.* 1978; Williamson *et al.* 1978; Young and Kuhar 1980). Furthermore, drugs that enhance GABAergic transmission do not generally appear to be anxiolytic (Sanger 1985). Indeed, 'It is surprising that a neurotransmitter receptor with such a ubiquitous distribution as the GABA-A receptor can serve as the target for therapeutically useful anxiolytic drugs with excellent tolerability. The explanation is that the mechanism of allosteric modulation by the benzodiazepine receptor does not profoundly alter the spatial and temporal pattern of GABAergic synaptic activity. Furthermore, potentiation of GABAergic transmission by benzodiazepine receptor ligands should be most effective in neuronal networks engaged in pathological overactivity' (Haefely 1992, pp. 165–6).

Given the important role that GABA plays as an inhibitory neurotransmitter, one would expect that a drug which facilitates its action would act as a general sedative rather than a specific anti-anxiety agent. Equally, given the known relationship between lowered GABA function and the occurrence of convulsions (Meldrum 1975), one might expect such a drug to be anticonvulsant. This appears to be true of both the benzodiazepines

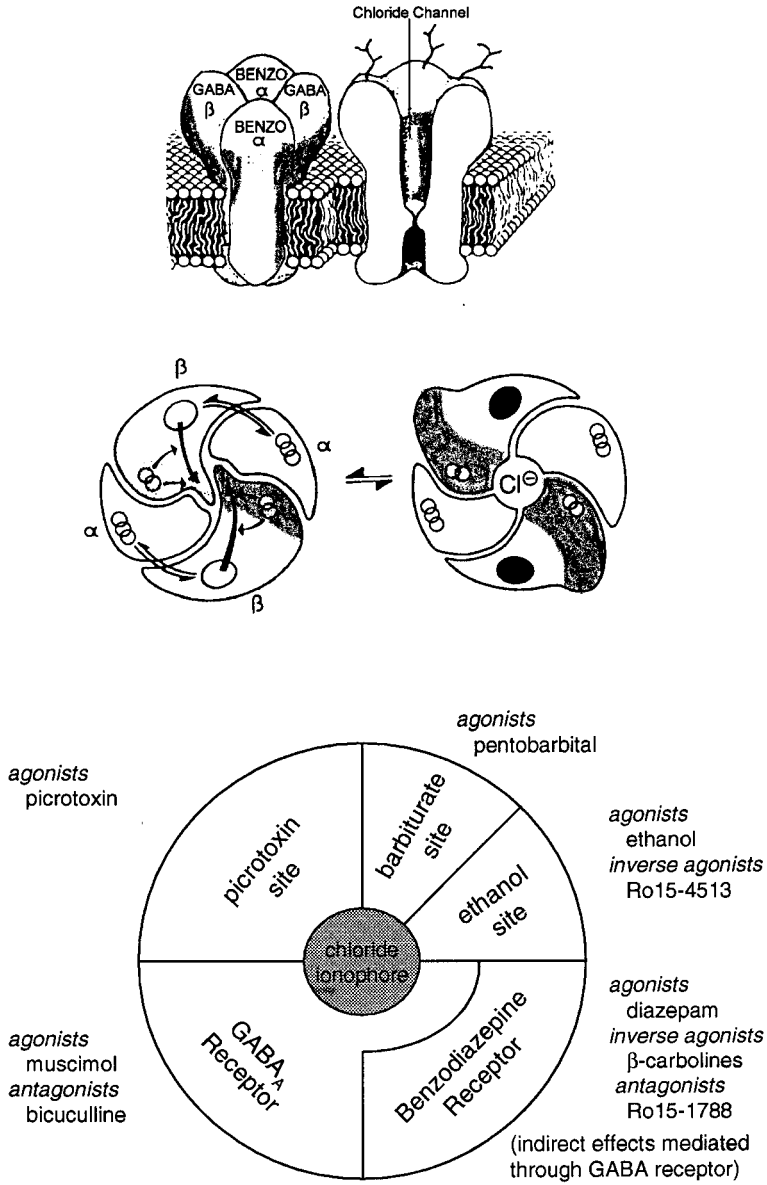


Fig. 4.1 The benzodiazepine receptor. This is linked to subunits of the $GABA_A$–chloride ionophore (top) and binding of GABA to its receptor causes a conformational change (middle) which causes the channel to open and chloride to pass through and so hyperpolarize the cell. (From Haefely 1990b.) The receptor complex also has distinct sites which can bind picrotoxin, barbiturates, and ethanol—all of which can affect the chloride channel directly (bottom) and so act like GABA. The benzodiazepine receptor by contrast (bottom, from Teicher 1988) interacts with the GABA receptor rather than the ionophore itself. It can thus increase (agonists) or decrease (inverse agonists) the effects of GABA. Benzodiazepines, therefore, only have effects if GABA is released and benzodiazepine antagonists are not GABA antagonists.

and barbiturates (Schallek *et al.* 1972; Browne and Perry 1973; Randall and Kappell 1973; Simon and Soubrié 1979). The question arises, are the apparent anti-anxiety actions of the classical anxiolytics secondary to their sedative, muscle relaxant, and perhaps anticonvulsant properties? We shall conclude below that this may be true of their immediate effects on anxiety symptoms, but not of their longer-term effectiveness.

Before leaving them, we should note one pharmacological peculiarity of the benzodiazepines. Benzodiazepine agonists appear straightforward enough. They act to increase the affinity of the GABA receptor and so enhance GABA action. However, as we remarked above, benzodiazepines act at a receptor which alters the affinity of the GABA receptor rather than at a receptor (such as the GABA receptor itself) which interacts with an ion channel. As a result, benzodiazepine antagonists do not block the effects of GABA. Rather, they block only the effects of any endogenous or exogenous benzodiazepine receptor ligand—which means that, in most behavioural tests, they have little or no effect until a benzodiazepine is administered. Thus the benzodiazepine antagonist flumazenil has little clinical action (Haefely *et al.* 1992), although it can reduce performance-induced anxiety, possibly by antagonizing the effects of an endogenously released (anxiogenic) benzodiazepine receptor ligand (Kapczinski *et al.* 1994). Particularly unusual, in the context of conventional pharmacology, are what are called benzodiazepine 'inverse agonists'. These act to reduce the affinity of the GABA receptor and so decrease the effects of GABA. They thus have effects opposite (hence 'inverse agonist') to those of the anxiolytic benzodiazepines.

Three final points should be noted about benzodiazepines. First is that the classification of a compound as agonist, antagonist, or inverse agonist depends to some degree on the behavioural test used (File and Cooper 1985). This is consistent with the more selective effects on anxiety of partial agonists. These are presumed to act as full agonists in some parts of the nervous system and as antagonists or even inverse agonists in other parts. The second point is that, as far as is known, the benzodiazepine receptor complex involves only $GABA_A$ and not $GABA_B$ receptors. The third is that there is now considerable evidence for the existence of endogenous benzodiazepines, which may be the natural ligands of the receptors (Median *et al.* 1993).

4.4 BUSPIRONE AND IPSAPIRONE

The treatment of anxiety with buspirone has occurred only in the decade since the first edition of this book. This development has revolutionized views of the general properties of anxiolytic drugs. It has also, as will be seen throughout this book, added considerably to our understanding of the neural mechanisms underlying anxiolytic action. The fact that it has not upset the septo-hippocampal emphasis of the first edition provides extremely strong support for this aspect of our theory. In a sense, the theory of the first edition can be said to have predicted the critical septo-hippocampal effects (both neural and behavioural) of buspirone without prior knowledge of the existence of this drug or, indeed, of any drug with a non-classical (Fig. 4.1) mode of action.

Except for its anxiolytic effects (Goldberg and Finnerty 1979), buspirone is so unlike the classical anxiolytics that it has been touted as an 'anxio-selective' drug (e.g. Goldberg

1990). This is overstating the case since, like all other drugs, it has a variety of unwanted and wanted side-effects. These include dizziness, nervousness, headache, nausea, insomnia, somnolence, and a range of less common symptoms.[1] Some of these may well result from the increased corticosterone levels produced by the drug (Urban *et al.* 1986). In addition to its anxiolytic action, and unlike the majority of benzodiazepines, buspirone is an antidepressant. However, buspirone (like other $5HT_{1A}$ agonists) is selective to anxiety in the sense of not sharing with classical anxiolytics their sedative, hypnotic, euphoriant, anticonvulsant, muscle relaxant, or addictive properties. It is superficially unlike the benzodiazepines, in that its effects on the symptoms of anxiety depend on long-term (10–14 days) administration (Feighner 1987). However, if measures such as the Hamilton Anxiety Scale are used, there appears to be 'no overall significant difference between buspirone and the benzodiazepines in treating patients with anxiety symptoms, with a maximum effect for both medications between the 2nd and 4th week of study' (Schuckit 1984, p. 64).

Buspirone appears to owe a large number of its actions, including probably its clinical efficacy, to interaction with the $5HT_{1A}$ receptor, with which the classical anxiolytics do not interact. Its antidepressant action is likely to be the result of a consequential down-regulation of $5HT_2$ receptors (Robinson 1993). It also interacts with dopamine D_2 receptors and indirectly affects noradrenergic systems (Jann 1988), something we return to later in the book. The fact that ipsapirone appears equally anxiolytic (see Glaser 1988) is consistent with the hypothesis that it is the $5HT_{1A}$ rather than dopamine D_2 actions of buspirone that are critical for its anxiolytic effects, since ipsapirone lacks the latter. However, the most important property of buspirone for the present argument is that it does not increase the effects of GABA nor interact directly with the GABA receptor. It has been reported to increase benzodiazepine binding *in vivo* (Geoders *et al.* 1988), but its behavioural effects cannot be blocked by benzodiazepine receptor antagonists (see Jann 1988). Thus, the reason that buspirone does not share the common side-effects of the classical anxiolytics is presumably because they do, but buspirone does not, interact with GABAergic transmission. Similarly, the presumption is that buspirone and classical anxiolytics are clinically effective in treating the symptoms of anxiety because they all act indirectly on the same final common path within the brain, but do so through quite distinct primary mechanisms. A major task of later chapters is to identify this final common path.

In many, but not all, animal tests buspirone has an inverted ∪-shaped dose–response curve. There is some indication (Johnston and File 1988) that this is because, at higher doses (above 1 mg/kg in rats), it releases corticosterone (Urban *et al.* 1986; N. McNaughton *et al.* 1996), which then acts, effectively, as an antagonist. The inverted ∪ dose–response curve may arise also because buspirone is only a partial rather than a full agonist at $5HT_{1A}$ receptors.

Some antidepressants are also apparently effective in the treatment of anxiety disorders. Where panic or obsessive–compulsive disorder is involved, the results cannot necessarily be taken at face value, since the primary effect of the drug may be on the panic attacks,

1. 'Adverse drug experiences reported following marketing of buspirone; period of report October 1984–February 1989', unpublished manuscript kindly supplied by Bristol-Myers.

obsessions, or compulsions with only a secondary, consequential, effect on reported anxiety (see Chapter 11). However, the tricyclic antidepressants are at least as effective as benzodiazepines in the treatment of generalized anxiety disorder. This might appear to make the distinction between anxiety and depression difficult to maintain, especially when we remember that buspirone is also an effective antidepressant. However, there are two reasons for believing that there are distinct neural systems which are involved in anxiety and depression to different extents. First, most benzodiazepines are much more effective in treating anxiety than depression. The specific exceptions, like alprazolam (see, for example, Dellemijn and Fields 1994, Table III), merely remind us that different drugs of the same chemical family can have different effect profiles. Second, there is no evidence that monoamine oxidase inhibitor antidepressants are effective in controlling generalized anxiety disorder. Where there are reports of apparently anxiolytic effects of the monoamine oxidase inhibitors, these appear to be in cases of atypical depression or anxiolytic-resistant anxiety with phobia (Sheehan *et al.* 1981), in which they are much more effective than tricyclics (Sheehan *et al.* 1981; Klein, 1993).

While considering such dissociations, we should note that buspirone, unlike imipramine, is completely ineffective in controlling panic (see Deakin 1993). This is despite their common action on generalized anxiety and unipolar depression and their common interaction with $5HT_{1A}$ systems.

As with buspirone, a task of later chapters will be to show how far the anxiolytic effects of antidepressants can be attributed to effects on final common paths in the nervous system. Given the assumption that buspirone reduces anxiety in virtue of its action as an agonist at $5HT_{1A}$ receptors, it is possible that both tricyclic and monoamine oxidase inhibitor antidepressants may vary in the ratio of their relative anxiolytic to antidepressant potency depending on the extent to which they are effective at $5HT_{1A}$ as opposed to other 5HT sites. An insufficient range of antidepressants has been tested on generalized anxiety disorder to assess this possibility.

4.5 CLOMIPRAMINE AND OBSESSIVE–COMPULSIVE DISORDER

Clomipramine is structurally very similar to standard tricyclic antidepressants like desipramine and imipramine. It is also antidepressant in its own right. However, it has recently been shown to be particularly effective in the treatment of obsessive–compulsive disorder (see Rapoport 1989; Modigh 1990). At first sight, this observation conflicts with the distinctions between anxiety and depression and between the treatments for each, since obsessive–compulsive disorder is of course classified among the anxiety disorders. However, the gulf between anxiety and depression is in fact widened by these findings, since 'the majority of investigators have come to the conclusion that the effect of clomipramine in obsessive–compulsive disorder is unrelated to its antidepressant effect' (Modigh 1990, p. 94). It is also possible in any case that, in obsessive–compulsive disorder, clomipramine has no direct effect on anxiety *per se*. Given its effectiveness in treating trichotillomania (see Rapoport 1989), and Rapoport's argument that obsessive–compulsive disorder reflects a disorder primarily of the basal ganglia, it may be that

clomipramine has a primary effect on obsessions and compulsions, and that it is the removal of these that eliminates the causes of anxiety. A further possibility is that clomipramine is at least partially effective in treating the anxiety symptoms associated with obsessive–compulsive disorder because of an action similar to that of imipramine at $5HT_{1A}$ sites, while being more effective than imipramine in the specific treatment of obsession because of a greater action on monoamine systems in the basal ganglia.

4.6 BETA-BLOCKERS

The 'beta-blockers' act at the β subclass of adrenergic receptor and affect both central and peripheral adrenergic transmission. Discussion is complicated by the fact that we here address central anxiolytic action, and yet the beta-blockers appear to act largely in the periphery; and also because one of the most commonly used drugs of this class, propranolol, is also a non-selective 5HT antagonist (see Griebel 1995, Table 1). However, propranolol too appears to have its greatest effect on somatic rather than psychic symptoms of anxiety, and to be as effective as benzodiazepines only in those patients whose predominant anxiety symptoms are somatic (Noyes *et al.* 1981). Clinically, beta-blockers are used most commonly for performance anxiety. Given these complications and the paucity of data on the behavioural effects of these drugs in animals, we leave them out of the discussion in the remainder of the book.

4.7 CLINICAL ACTIONS OF THE ANXIOLYTICS—A SUMMARY

The data we have briefly reviewed so far suggest that there could be a single neural system which mediates the common effects of all the anxiolytic drugs. However, the anxiety affected by these drugs must be distinguished from atypical depression, simple phobia, panic, and obsession. For, in the case of each of these conditions, there are drugs which are ineffective in treating them but effective in treating generalized anxiety (Table 4.1). For example, while atypical depression can be treated with the antidepressant phenelzine, it does not respond to benzodiazepines. Phenelzine offers a particularly useful point of comparison in Table 4.1, since atypical depression shares many, particularly autonomic, symptoms with those of the anxiety disorders.

All drugs have multiple effects. The effects that are therapeutically undesirable or irrelevant (in any particular context) are termed 'side-effects'. From a physiological point of view, of course, there is no reason to suppose the side-effects to be any less closely related to the drug's principal mode of action than is its therapeutic effect. As we have noted, the benzodiazepines, for example, usually act not only as anxiolytics but also as muscle relaxants and anticonvulsants. If we look only at these compounds, we cannot tell whether these additional actions are crucial to their anxiolytic effects or totally unrelated to anxiolytic action. This is true even of the partial agonists, whose relative lack of muscle relaxant and addictive effects appears to be the result of an incapacity to produce sufficient functional receptor occupancy. The side-effects of the benzodiazepines are shared

Table 4.1 The effects of various classes of putatively anxiolytic drugs on anxiety syndromes and their actions on receptors in the brain in comparison to the antidepressant monoamine oxidase inhibitor, phenelzine. Phenelzine has been included in the table because, like buspirone and imipramine, it is effective in treating depression and, in addition, appears to be particularly effective in treating atypical depression—a form of depression in which many of the symptoms overlap with those of the anxiety disorders. (For details see Appendix 1)

	BDZ	BUS	IMI	CMI	PHEN
Phobia	0	?	?	?	?
Generalized anxiety	–	–	–	–	0?
Panic attacks	(0)*	0	–	–	–
Obsessions/Compulsions	0	?	(–)	–	(–)
Atypical depression	0	?	(–)	?	–
Unipolar depression	0*	–	–	–	–
Relaxant, anticonvulsant	+	–	0	0	0
BDZ/GABA	+	0	0	0	0
$5HT_{1A}$	0	+	+	?	?
5HT (uptake/breakdown)	0	0	–	–	–

*Excluding alprazolam (e.g. Sanderson *et al.* 1994).
BDZ, benzodiazepine; BUS, buspirone; IMI, imipramine; CMI, clomipramine; PHEN, phenelzine; GABA, γ-aminobutyric acid; $5HT_{1A}$, 5-hydroxytryptamine 1A receptor; 0, no effect; –, reduction; +, increase; (), small or discrepant effects.

by all other classical anxiolytic drugs. At the time of the first edition, therefore, muscle relaxant or anticonvulsant action might have been thought necessary properties for an effective anxiolytic. Now, in contrast, whatever the relationship between sedative, anticonvulsant, and anxiolytic effects in the classical anxiolytics, there is incontrovertible evidence that an anxiolytic need not be sedative or anticonvulsant: buspirone, ipsapirone, and imipramine are anxiolytic drugs which are as effective clinically as benzodiazepines, but share none of their side-effects.

This clinical effectiveness is subject to one apparent caveat. The novel anxiolytics require at least 10 days of administration to produce their therapeutic effect. In this respect, these compounds are generally seen to differ markedly from the benzodiazepines, which are believed to be immediately effective and preferred under some circumstances for this reason (Trimble 1990, p. 15). However, as measured by the Hamilton Anxiety Rating Scale, benzodiazepines take as long as novel anxiolytics to produce their effects (e.g. Wheatley 1990, Fig. 14.2). Even in the case of pre-operative anxiety, which might have been expected to be especially susceptible to acute anxiolytic effects, the benzodiazepines do not act immediately (Wikinski *et al.* 1994). Given these data, it appears possible that the benzodiazepines have two separate effects. First, there are immediate euphoriant and muscle relaxant actions, which are known to show rapid tolerance; and second, there is a truly anxiolytic effect which develops over time and thus appears, coincidentally, as the muscle relaxant and euphoriant effects disappear. The anxiolytic effect itself is one of the few that do not show tolerance. For example, patients treated with long-term alprazolam for panic disorder show apparently normal baselines, but greatly depressed responsiveness to diazepam in tests of saccadic eye movements, growth

hormone secretion, memory, and self-rated levels of sedation. By contrast they show no differences in diazepam-induced changes in measures of anxiety or cortisol levels (Cowley *et al.* 1995).

The lack of specificity to any one DSM-IIIR-defined type of anxiety of this wide range of anxiolytic drugs is consistent with the unitary view of the behavioural inhibition system presented in Chapter 3. The fact that, taken as a class, the drugs do not affect simple object phobia or panic (see Chapter 11) is consistent with the separation of the behavioural inhibition system from the fight–flight system. In addition, the consistency of the effects of both novel and classical anxiolytics forms the basis for the neuropsychology of anxiety presented in this book. However, a complete neuropsychology must also account for the differences between the subclassifications of anxiety (a topic dealt with in Chapter 11), as well as for the clinical experience that nonetheless leads to their being grouped together (a major theme of Chapters 12 and 13), as for example in the DSM system.

4.8 BRIDGING THE SPECIES GAP

Despite the widespread use of DSM-IIIR and related schemes, there is little agreement among psychiatrists as to exactly what anxiety is. However, as we have noted, there is general agreement as to which drugs are anxiolytic, and this agreement has been good enough for us to use the drugs to question some details of the DSM-IIIR classification (see Table 4.1; Chapter 11). As we have also seen, classical and novel anxiolytics have quite different primary actions on the brain. This gives us two reasons for reviewing the effects of the anxiolytic drugs in animals. First, the behavioural effects common to both classical and novel anxiolytic drugs will represent a profile of truly anxiolytic, as opposed to other, action. From this profile we shall later obtain some insights into the psychological nature of the effects of the drugs. Second, the neural effects common to these drugs should indicate the neurological basis of their anxiolytic action. By comparing the behavioural profile obtained in the present chapter with the effects of lesions of the nervous system we shall, in the following chapters, narrow the search for the neuropsychology of anxiety.

However, at this point we must make explicit an assumption that has till now been implicit only: that anxiety exists in other animals as well as human beings, and that they both respond in much the same way to anti-anxiety drugs. This assumption is critical, not only for our arguments but also for the whole research endeavour on which they are based. For it is impossible to perform the great majority of the relevant experiments without using animal subjects. Yet many people will find this assumption hard to accept. Anxiety is commonly believed to be a uniquely human state, dependent on such complex cognitive capacities as the ability to foresee the future, to form a self-image, or to imagine one's own mortality. To the extent that this belief is correct, the present approach to the study of anxiety is misconceived. Nor is it a belief that can be lightly brushed aside. The general continuity of human and animal behaviour is not open to serious dispute on biological grounds. But this does not avoid the need to demonstrate that there is sufficient continuity in each specific case.

Indeed, a recent criticism of the theory of the first edition of this book (M. W. Eysenck 1992b, p. 166) was that:

> Gray may well be right that there are important similarities between rat and man in terms of the brain structures which mediate anxiety, but it is much less likely that the cognitive processes associated with anxiety are the same in the two species. . . . Even if the layman or clinician were to accept that perceiving events as, say, signals of punishment or non-reward called forth biological responses that we have in common with other species, they might still argue that the cause of complaints of anxiety was in perceiving events in this way, and not in possessing the biological mechanisms of these responses.

Eysenck argued, in particular, that the theory of the first edition failed to capture some recently discovered, specifically cognitive, components of human anxiety because of its foundations in animal work. (We discuss these cognitive components in some detail in Chapter 13.)

In one sense, of course, human anxiety cannot be like that of any other species. We have our own unique sensory and cognitive capacities and this must affect the way we perceive stimuli able to elicit anxiety. Yet, taken to the extreme, this is a counsel of despair that would imply that we cannot generalize even from one human being to another, since everyone has his or her unique concerns. However, if anxiety and other emotions are viewed as resulting from particular functional classes of situation that are species-general (McNaughton 1989a), but which may produce certain species-specific adaptations, then the similarities between species and individuals would be expected vastly to outweigh the differences. To take a physiological example, human beings neither eat grass nor ruminate, and cows do not eat meat. Nonetheless there are virtually no differences in the basic physiological principles of digestion between the two species, and surprisingly few detailed practical differences. In the same way, we accept Eysenck's point that the specific stimuli which give rise to human anxiety on any particular occasion can be unique, and that their anxiogenic potential depends upon the interpretation placed upon them by the individual. The cognitive processes that give rise to such interpretations share with other animal species a reliance upon such universal principles of learning as those of classical conditioning; but undoubtedly they also depend upon specifically human cognitive processes that cannot occur in a non-linguistic species. However, once formed as anxiogenic stimuli, their action in eliciting the emotion of anxiety is essentially similar in man and other mammalian species. Thus, a full analysis of anxiety requires an understanding of both the specifically human cognitive processes that lead to a particular individual's construing a particular event as a threat, e.g. to self-esteem (the kind of process emphasized by Eysenck), and the processes common to the human and other species which convert that construal into the emotion of anxiety. (For a recent wide-ranging treatment of the former area, the reader is referred to Dalgleish and Power, 1999.)

As we shall now see, the observed effects of anxiolytic drugs in animals suggest that they act on an anxiety which is very similar to the human kind. The next crucial step will be taken in Chapter 11. There we expound a theory of anxiety based largely on animal experiments but apply it to the phenomena of anxiety in man. In the last analysis, this is the touchstone of the theory's success: the understanding it brings, or fails to bring, of human anxiety.

4.9 THE BEHAVIOURAL EFFECTS OF ANXIOLYTIC DRUGS IN ANIMALS

In what follows, we progress from a specific to an increasingly general view of anxiolytic drug action, and hence anxiety. In Chapter 2, we took as our paradigmatic case the behaviour that is likely to occur in the context of potential threat. But we concluded that the critical feature for such behaviour is the approach–avoidance conflict inherent in a situation of potential threat, rather than the potentiality of the threat as such. In the present chapter we deal with explicit cases of approach–avoidance and find that, as with the case of predatory defence, the anxiolytic drugs shift the balance of the conflict from avoidance towards approach. In Chapter 3, we found reason to broaden the notion of approach–avoidance conflict and to treat as potentially threatening: novel stimuli of certain kinds, signals of punishment, and signals of reward omission, in addition to the (ethologically speaking) paradigmatic innate anxiety stimuli. In what follows we shall deal with all of these cases, and find reason to emphasize both the role in anxiety of passive (inhibitory) avoidance and the necessity for true conflict between highly primed (activated) incompatible responses. This leads naturally to the fact that anxiolytic drugs also shift an active avoidance–passive avoidance conflict away from inhibitory avoidance. At the end of the chapter we shall find reason to generalize still further. The drugs are effective in a number of tasks which appear, at first sight, to have no aversive component at all, and instead appear to involve memory. However, the tasks do involve conflict—but now a conflict between approach and approach. Even here, though, a common thread can be seen: correct performance of the task requires inhibition of competing responses that are currently inappropriate.

A careful description of the effects of the anxiolytic drugs on the behaviour of experimental animals, and where possible of human beings, is essential, then, for the arguments pursued in this book. There have been many hundreds of experiments on these effects, starting with the pioneering work of Jules Masserman on alcohol and Neal Miller on sodium amylobarbitone (Masserman and Yum 1946; Miller 1951). In Appendix 1 we summarize and update the conclusions of earlier reviews, especially of the effects of the classical anxiolytics. An overall summary of summaries is presented in Table 4.2. Importantly, the behavioural effects of the anxiolytic drugs are in most cases independent of the drug used or the species investigated. The following discussion therefore refers equally to the benzodiazepines, barbiturates, and alcohol. The effects of meprobamate have not been investigated to anything like the same extent, but are generally similar to those of the other classical anxiolytics.

In the context of the remainder of the book, a crucial question is how far the behavioural effects of the classical anxiolytics are reproduced by novel anxiolytics such as buspirone. As will be seen, where buspirone has been investigated, the similarities are surprisingly good, extending to some tasks which are superficially unexpected (such as the Morris water maze). However, it should be remembered that buspirone has seldom been tested more than once with any particular experimental régime (with the exception of the classic 'screening tests' for anxiolytic action), and its entries in Table 4.2 usually reflect only one experimental régime for each major class of behavioural test. Where null results are encountered in the literature on buspirone, it should also be borne

in mind that it usually has an inverted ∪ dose–response curve and can have more convincing anxiolytic effects with chronic pretreatment (e.g. Zhu and McNaughton 1995), possibly because of tolerance of dopaminergic action or tolerance of a corticosterone-based antagonism (see Matheson *et al.* 1996; and the earlier discussion above).

Throughout what follows, we consider only the effects of low doses of the drugs. In the case of the classical anxiolytics, these are doses which do not produce sedative, motor, or anaesthetic effects (e.g. 5 mg/kg of chlordiazepoxide in the rat). In the case of buspirone, they are doses which, across a range of tasks, produce effects quantitatively similar to those of low doses of benzodiazepines, and are below those which show diminished or reversed effects (usually less than 1 mg/kg).

In looking at the behavioural effects of the anxiolytic drugs we shall take the opposite approach to that adopted by most modern screening tests. These are designed for speedy testing of many compounds, but allow minimal theoretical analysis and appear each to be measuring a somewhat different behavioural trait with no common factor of anxiety uniting them (File 1992; see also Broekkamp *et al.* 1989; Davies *et al.* 1994; Dawson *et al.* 1995). The problems which result from this approach have led to more ethological analyses (e.g. Cole and Rodgers 1993; Rodgers and Cole 1994; Miczek *et al.* 1995), but not usually to the parametric analysis which, in the Blanchards' hands, showed the ∪-shaped relationship of many behavioural measures to the underlying construct of anxiety (Chapter 2; see also Rodgers and Cole 1993). One exception to this is the elevated T-maze test developed by Graeff's group (Viana *et al.* 1994); but the advantages of this test stem from its closer relation to the older (and more laborious) operant analysis of drug effects. As noted by Dawson and Tricklebank (1995), the increased labour of ethological analysis removes the time-saving which is perhaps the sole attraction of the conventional screening tests. We shall also totally ignore the frenzy of pharmacological analysis which has been directed to the conventional screening tests, since this has usually generated more heat than light (see, for example, Griebel 1995). In contrast to the screening tradition, our approach is to treat the entire literature as one massive test battery and to interpret the results in terms of our prior ethological and learning-theory analysis. Thus, as we have done in miniature already in the last two chapters, we shall use a lack of effect in one test to rule out the various psychological processes involved in that test from any contribution to the effect of the drugs obtained in some other test.

Since many experiments require the use of appetitive or aversive stimuli to control behaviour, we start our review with the simplest effects of these stimuli. Our immediate focus will be on ruling out effects of the drugs on active approach and active avoidance so as to allow a clear interpretation of the effects, described later, on the inhibitory avoidance inherent in approach–avoidance conflict.

In Chapter 2 we concluded that the paradigmatic case of pure fear, the response to an immediately present predator, is unaffected by anxiolytic drugs. This result encompassed each of fight, flight, and freezing. Consistent with this conclusion, in more artificial laboratory situations this form of behaviour is again not affected. A key finding here is that, with simple active avoidance (as opposed to two-way avoidance, which we consider shortly), there is no effect of the drugs, matching their lack of effect on the Blanchards' Fear Defence Battery. It can also be seen from Table 4.2 that anxiolytic drugs do not

Table 4.2 Comparison of the common behavioural effects (ANX) of novel (NOV) and classical (CLAS) anxiolytics with the common effects (S/H) of septal (SEP) and hippocampal (HIP) lesions and lesions of the amygdala (AMYG) dorsal ascending noradrenergic bundle (NA), ascending serotonergic system (5HT) and blockade of the cholinergic system (ACh) (for details see appendix 8 on http://www.oup.co.uk/neuropsychol_anxiety).

	NOV	CLAS	ANX	S/H	SEP	HIP	AMYG	NA	5HT	ACh
Eating	+	+	+	0	0/+	0	0	(0)	+	
Drinking	0	+	0	0	+	0	0	(0)	0	
Response to aversive stimulation	–	0	0	+/0	+	+/0	–	(0)	0/+	
Aggression (no shock)		+	+	–	–	–	–	0		–
Escape	0	0	0	0	0	0	–	0	0	
Simple rewarded learning	0/–	0	0	0	0	0	0	0	0	–
Frustration		0	0	0	0	0	–	0		
One-way avoidance	0/–	0	0	0	–	0	–	0		0
Classical aversive conditioning		0	0	0	0	0	–			
Conditioned suppression on baseline	–	–	–	0	(–)	0	–	–	–	
Conditioned freezing		–	–	–	–	–	–	+/0		0
Defensive burying	–	–	–	0/–	–	0	0			
Fear-potentiated startle	–	–	–	0		0	–		+	
Passive avoidance	–/0	–	–	–	–	–	–	0/–	–	–
Two-way avoidance		+	+	+	+	+	–	0/+		+
Non-spatial avoidance		+	+	+	+	+		0		+
Extinction		–	–	–	–	–	0	–	–/0	–
Reversal		(–)	(–)	–	–	–	0/–	0		
Successive discrimination	–/0	–	–	–	–/0	–	–	–	0	–
Single alternation		–	–	?	+?	+/0/–	–	–		
Simultaneous spatial discrimination		0	0	0	0	0		0		
Spontaneous alternation		–	–	–	–	–	–	–		–
Radial arm maze		0/–	0/–	–	–	–	–			–
Water maze	–	–	–	–	–	–	0	0		–
Defecation	+	0	0	0	–	0		0		

Rearing	–	–	–	–	–	–	+	–	–	–
Social interaction	–/0	–	–	–?	–?		–	0	–	
Elevated + maze	–/0	–	–	–	–		0		–	+
PREE		–	–	–	–	(–)	0	–		
Conditional/delayed discrimination	(–)	(–)	(–)	(–)		–/0	0		+	–
Latent inhibition				–	–	–	0	0		
DRL	–/0	–	–	–	–	–	–	0	–	–
Fixed interval	–/0	–	–	–	–	(–)	0	0/–		0

PREE, partial reinforcement extinction effect; DRL, differential reinforcement of low rates of response; +, improved/increased; –, impaired/decreased; 0, no change; mixed symbols, different results in different experiments; symbols in brackets, questionable changes.

affect any responses related to simple appetitive motivation, nor to the learning of active responses to avoid noxious stimuli or approach appetitive ones. These results immediately rule out any general actions of the anxiolytics on simple motivation (including pure fear), on sensory discrimination, on motor control, or on learning and memory in general.

As can be seen from Table 4.2, in tests in which there is explicit approach–avoidance conflict (Miller 1951) involving a painful stimulus, such as foot-shock, as was the case with predatory defence (Chapter 2), the anxiolytic drugs shift the balance of the conflict from avoidance towards approach. As previously noted in Chapter 2, the effect of the drugs on risk assessment behaviour is a complex function of defensive distance. This function explains, for example, the opposite effects of anxiolytics on rearing in open field tests depending on whether loud noise and bright lights are used or not.

The same pattern of reduced passive avoidance is seen when the avoidance component of the approach–avoidance conflict is based not upon a painful stimulus, but upon non-reward. Perhaps the clearest cases of this are extinction and successive discrimination. Anxiolytic drugs have no effect, in the rat, on learning to run down a runway, but they greatly impair learning not to run when the food reward is no longer available. Similarly, when responding is reinforced by food in the presence of one stimulus (S+) and a second stimulus (S–) explicitly indicates that food is not available (or that shock will be presented), the animals can learn the required response if the S+ and S– are presented as a concurrent choice (simultaneous discrimination) but not if they are presented successively, so that the S– requires the animal to inhibit the responding which obtained reward in the immediately preceding S+ period.

The selective action of anxiolytics against passive avoidance is seen also in tasks in which the conflict is with an active avoidance response, as in the well-studied two-way active avoidance task in the shuttlebox. The effect of the drugs in this task is to improve active avoidance, consequent upon a reduced tendency to inhibit returning to that portion of the apparatus from which the animal has just escaped.

The actions of the anxiolytics on the bulk of operant schedules (fixed interval, FI, differential reinforcement of low rates of response, DRL, etc.) are again consistent with the notion that they reduce inhibition of responses that are followed by both reward and non-reward. In all cases, the way in which responding has changed indicates that the animal retains knowledge of the task to be performed, but that its capacity to inhibit prepotent responding is reduced.

The data on anxiolytic drug action in Table 4.2 thus license the generalization from behaviour which occurs in the context of approach to potential predators to behaviour which involves approach to a wide array of conditioned and unconditioned aversive stimuli. Additional data in the table show that we must generalize still further. The drugs are effective in a number of tasks which appear to have no aversive component at all, involving instead memory. However, this generalization does not involve as big a step as might at first appear: the tasks do involve conflict, but conflict between approach and approach. In delayed matching to sample with trial non-unique stimuli, for example, in order to approach the correct goal on any particular trial, the animal must inhibit responding to a goal which would have been correct previously (often on the immediately preceding trial). Thus, correct performance of the task requires inhibition of competing responses that are currently inappropriate.

Probably the most theoretically significant of these tasks is the Morris water maze (Morris 1981, 1984). This is the quintessential test of spatial navigation, and nowadays the most commonly used test of temporal lobe dysfunction in non-human subjects. It consists of a circular featureless swimming pool containing a submerged, and hence invisible, platform which is always located in the same spatial position. The rat is placed in the pool at different positions on different trials. Control rats quickly learn to find the platform and to swim almost directly to it. The invisibility of the platform also allows the use of transfer tests to determine what the rat has learnt. In such tests, at the end of acquisition, the rat is placed in the water as usual, but there is no platform to swim to. Control rats show that they know the precise position at which the platform should have been located. They swim to this position and then swim in very tight circles on the spot where the platform would have been (Fig. 4.2). Chlordiazepoxide, buspirone, the $5HT_{1A}$ agonist 8OH-DPAT, ethanol, and other anxiolytic drugs eliminate acquisition of spatial navigation in the water

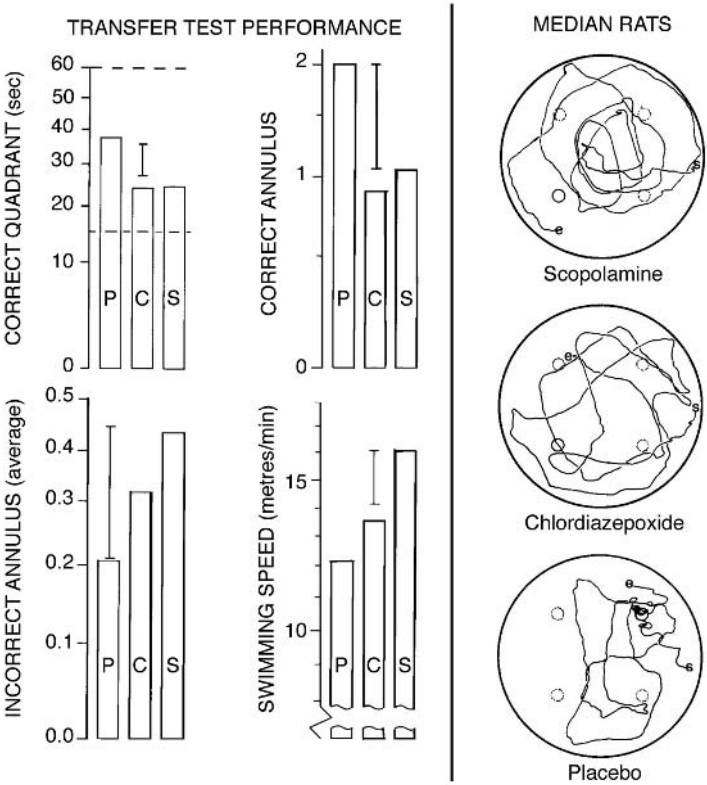


Fig. 4.2 Effects of the benzodiazepine chlordiazepoxide and the anticholinergic scopolamine in a transfer test after spatial learning in the water maze. Both drugs impair the capacity of the rats to find the location at which a platform had been during learning. Note that the paths taken by the median rats of each group show that control rats tend to swim with a zigzag path to where the platform was and then circle on the spot, that scopolamine rats tend to swim in circles, and that chlordiazepoxide rats show no consistent pattern. P, placebo; C, chlordiazepoxide; S, scopolamine. (From McNaughton and Morris 1987.)

maze (McNaughton and Morris 1987, 1992; Devenport *et al.* 1989; Rowan *et al.* 1990; McNamara and Skelton 1991, 1992; Carli and Samanin 1992), while leaving intact non-spatial (or less accurate) strategies for finding the platform. It is of note that in this task, unlike many of the standard screening tests for anxiolytics, buspirone has a purely linear dose–response curve (McNaughton and Morris 1992).

There are two features of the effects of anxiolytics in the water maze which are particularly important for the analysis of behavioural inhibition. These are discussed more fully in Chapter 8, but will be mentioned here. First, it is clear that the drugged rats are motivated to learn the task and are capable of at least some learning. On successive trials in the apparatus, they reduce the distance they swim and towards the end of acquisition appear, on the basis of path length, to be solving the task nearly as well as controls. However, in the transfer test, they swim straight through the position at which the platform would have been (Fig. 4.2) and show little indication that they know where it is. This suggests that they have a simple rule which allows them to bump into the platform when it is there. Second, while this task involves aversive motivation (deriving from immersion in water), it is formally an escape task, not one of passive avoidance. However, we concluded above that it is only passive avoidance, not escape or active avoidance, which anxiolytic drugs normally impair. The deficit produced by anxiolytic drugs in this task cannot therefore be attributed to the use of a negative reinforcer. Nor can it be attributed to a general failure of memory, since respectable learning occurs, but accuracy is reduced by a failure to determine the position of (or at least to correctly approach) the platform. Finally, the deficit cannot be attributed, except ad hoc, to a loss of specifically spatial capacities, since many of the other tasks we have considered (FI, DRL, etc.) are not distinguished by specifically spatial requirements. Further, administration of buspirone only during the probe trial at the end of acquisition has no effect (Rowan *et al.* 1990) and diazepam has no effect on performance (McNamara and Skelton 1991; see also Kant *et al.* 1996) or even on acquisition if the rats have first been habituated to swimming in the maze (Zanotti *et al.* 1994). This latter result is particularly important in relation to recent data (see below) which show that animals with hippocampal lesions can learn spatial navigation provided interference from prepotent competing strategies is eliminated. The latter factor, then, appears to be what unites the effects in the Morris water maze of both anxiolytic drug treatment and hippocampal lesions.

In a radial-arm maze task, the rat is placed into the centre of a maze in the shape of a star burst with, typically, 8 or 16 arms. In the simplest version of this task all arms contain a piece of food at the beginning of any one trial, and the most efficient performance is for the rat to visit each arm only once. This might seem very simple in that this outcome can be achieved by a rule such as 'turn sharp left as you come out of each arm'. In practice, rats do not solve the task in this way, tending rather to choose arms roughly opposite to that which they have just visited. They must, therefore, maintain in working memory information (presumably, spatial information) about the arms they have visited. We discuss different types of memory more in Chapter 8. Suffice it to say here that a normally anxiolytic dose of chlordiazepoxide (5 mg/kg) is without effect in this task (data from J. N. P. Rawlins cited by McNaughton *et al.* 1980b, p. 188), but a high dose (20 mg/kg) is effective (Rawlins, personal communication). The $5HT_{1A}$ agonists also impair performance in the radial-arm maze (Winter and Petti 1987).

When we view spatial tasks as a whole, we see a progression from simple to complex tasks. In the simplest, position learning in a T-maze, there is little effect of the drugs; in more complex tests (spontaneous alternation and radial-arm maze), there is a clear effect only at high doses; finally, in the most complex (spatial navigation as opposed to other solutions of the water maze), there appeared to be total abolition of the capacity for spatial learning. The importance of complexity, as opposed to space, is underscored by the fact that, if the radial-arm maze is modified to be more complex but less spatial, anxiolytics become effective even at low doses. Instead of putting food in all eight arms of the maze, one can place food in only four, and signal which four by placing non-spatial cues, e.g. a sandpaper floor, in each baited arm. The arms containing the food are changed from trial to trial. Chlordiazepoxide significantly impairs acquisition of this task at a dose of only 5 mg/kg (Hodges and Green 1986; McNaughton and Olaman, in preparation).

'Complexity' is an elastic term; it may relate well, however, to the concept of conflict that has been at the root of our analysis so far. In the case of the water maze, we can make some progress in drawing these ideas closer together. As we discuss in Appendix 8, septo-hippocampal lesions produce an even larger deficit in spatial navigation in the water maze than do anxiolytic drugs. However, there are a number of data that suggest that space *qua* space is not the critical factor in determining the lesion deficit. For example, Eichenbaum *et al.* (1990) found that rats with fimbria-fornix lesions (which disconnect the septum from the hippocampus) could perform the task nearly as well as controls, but only if they started from the same position on every trial; yet a probe trial showed that they were solving the problem using spatial cues and not some simpler rule (see also Whishaw and Tomie 1997). The main cause of the drug-induced deficit in the standard form of the maze may, therefore, be due to an inability to inhibit conflicting response tendencies when starting from different positions on each trial. This seems particularly likely, given the findings of Zanotti *et al.* (1994) that prior experience of the maze (which would be expected to eliminate initial exploratory tendencies) eliminates the effects of anxiolytics on spatial navigation. The concept of complexity may perhaps, therefore, be made more precise along these lines.

4.10 ANXIOLYTIC ACTION—SOME CONCLUSIONS

We have now completed a general overview of the behavioural profile of the anxiolytic drugs. We find little evidence that the drugs affect basic motivational, perceptual, or simple memorial processes. With only a few exceptions (for some of which we have been able to provide possible, albeit ad hoc, explanations), the effects of the drugs have been seen in tests which involve approach–avoidance conflict (with which we started our analysis in Chapters 2 and 3), or avoidance–avoidance conflict (as in two-way active avoidance), or approach–approach conflict (as in delayed matching to sample). In all these cases, whether because of reinforcement of specific responses or of the nature of the stimuli presented, the animal is faced with conflicting response tendencies of similar strengths. This is particularly a feature of the standard spatial tasks. The effect of the drugs is to reduce the power of the inhibitory component of the conflict (i.e. with approach–avoidance, they reduce avoidance and hence increase approach).

The extent to which this pattern is generally applicable to all of the chemically different anxiolytic drugs is remarkable, especially so in the case of buspirone (at least with low doses, lengthy administration, or block of corticosterone release). As noted earlier, buspirone shares none of the side-effects of the classical anxiolytics. It might have been expected, therefore, that a large part of the common behavioural profile of the classical anxiolytics would not be reproduced by this compound.

The clearest effects of buspirone and the clearest similarity to the effects of anxiolytic drugs are in what might be viewed as the least emotional, most stimulus-oriented measures: spatial navigation in the water maze and rearing in the low-stress open field. The least clear effects and the greatest departure from those of classical anxiolytics are seen, paradoxically, in tasks supposedly designed as screens for anxiolytic drugs: the elevated plus maze and social interaction. It may be that these latter tests suffer from being more tests of action at GABA–benzodiazepine receptors than of anxiety. But, as we have noted, another possibility is that the anxiolytic effect of buspirone is masked by an antagonist action of corticosterone, released in consequence of the inherently stressful nature of this type of test, which may be lost with repeated administration (as first suggested by Johnston and File 1988; see also Zhu and McNaughton 1995; Matheson *et al.* 1996; N. McNaughton *et al.* 1996). For our present purposes this issue is not critical, since in the great majority of cases benzodiazepine-like effects of buspirone can be obtained if suitable dosing régimes are used. Below, and in Appendix 10, we discuss the complex involvement of 5HT systems in both the regulation of the behavioural inhibition system and the control of other aversive behaviour. As summarized by Handley *et al.* (1993), the evidence 'that decreasing serotonergic function is anxiolytic and increasing it is anxiogenic, is most consistent in models of behavioural inhibition where the stimulus inhibits an approach response (conflict models). However, paradoxical drug effects are also frequent, especially where the aversive stimulus evokes an active response.'

The similarities between the anxiolytic drugs, classical and novel, strongly imply that they act upon a common neural system (or systems). Central to our understanding of the effects of the drugs is that they do not affect either the eliciting effects of rewarding or punishing stimuli, nor the learning of simple responses, nor the learning of discriminations reinforced by such stimuli. They also appear not to affect the simple eliciting properties of novel stimuli. Thus they do *not* act on systems mediating perception, motivation, action, or any general kind of reinforcement or memory. By far the widest range of their effects, and the most easily understood, are found in tasks where behavioural inhibition of some prepotent response is central to correct acquisition: passive avoidance, extinction, successive discrimination, etc. It appears, then, that anxiolytic drugs may be characterized as producing a more or less specific interference with the behavioural inhibition system, as defined by Gray (1975), the key properties of which have been detailed in Chapter 3.

Note that the behavioural inhibition system, so defined, does not mediate all cases where behaviour occurs at a low level and so might in some sense be described as inhibited. We have already seen this in the case of innate freezing. The stimuli which activate the behavioural inhibition system are taken to be secondary punishing stimuli, secondary frustrative stimuli, and other stimuli which have been paired with primary negative reinforcers, and also novelty and other innate anxiety stimuli. As argued in Chapter 3, the

common function of all such stimuli is to signal potential threat, and they activate the behavioural inhibition system only when they do in fact signal potential threat in such a way as to generate conflict. It follows that there can be both cases in which the frequency of behaviour is reduced by some process other than behavioural inhibition, and cases in which stimuli that initially produced their behavioural effects through the behavioural inhibition system subsequently cease to do so. An example of the former case is positive conditioned suppression, in which an animal can obtain a food reward by holding still; this is not affected by anxiolytic drug treatment (Miczek 1973). An example of the latter case is over-learned successive discrimination (S– signalling shock for responding); once the habit is learned to the point that shocks are no longer received, anxiolytics no longer disinhibit responding (McNaughton 1985c).

Nor is the behavioural inhibition system, as we define it, solely concerned to produce behavioural inhibition. As described in Chapters 2 and 3, its outputs include increased risk assessment behaviour and increased levels of arousal. Descriptively, these outputs involve behavioural activation rather than inhibition, yet they are sensitive to anxiolytic drugs (although the ∪-shaped relationship of risk assessment measures to level of anxiety discussed in Chapter 2 creates great problems for interpretation unless anxiety is parametrically varied within a single experimental paradigm).

The overt exploratory behaviour that forms part of risk assessment results in increased attention to environmental stimuli. The effects of anxiolytic drugs on this attentional output of the behavioural inhibition system have been studied in connection only with novelty as the input to the system. A lone experiment by McGonigle *et al.* (1967) has taken this type of analysis furthest, by measuring the degree to which a drugged animal learns about a novel element in its environment. These workers trained rats on a random 50 per cent partial reinforcement schedule of reward for choosing the positive cue (black or white) in a choice box, choice of the negative cue never being rewarded. The rats were then shifted to a combined-cue discrimination (black vs. white and horizontal vs. vertical stripes, both cues being presented together) in which the old positive cue remained positive. During this stage of the experiment half the animals received a barbiturate and half placebo. Finally, transfer tests were conducted with only horizontal vs. vertical stripes and no drug. Controls chose correctly on the transfer trials, but the drugged animals did not. Since the task was a simultaneous discrimination, not normally affected by anxiolytic drugs, it is likely that the critical element that made it sensitive to these agents was the requirement to attend, and transfer control, to a novel environmental cue. However, as already noted in the 1982 edition of this book, the problem of attention requires much more research before we can be sure whether anxiolytic drugs decrease this output from the behavioural inhibition system.

There are several examples of a blockade by the anxiolytic drugs of the increment in level of arousal which is postulated to result from activity in the behavioural inhibition system. This too remains a neglected problem. In part, this is because the concept of arousal is much less firmly embedded in animal learning theory than in human experimental psychology (e.g. Gray 1964; Eysenck 1967; Broadbent 1971), so that there are few experimental paradigms which unambiguously address this issue. One paradigm which comes close to this ideal is the partial reinforcement acquisition effect. This consists in an increase in running speed, relative to animals rewarded on every trial, in animals

rewarded on a 50 per cent random partial reinforcement schedule for traversing a straight alley. The partial reinforcement acquisition effect has been explicitly treated as reflecting an increased level of arousal arising from exposure to secondary frustrative stimuli (Amsel 1962, 1992; Gray and Smith 1969), and it is blocked by anxiolytic drugs. A second experimental demonstration of the increased arousal output of the behavioural inhibition system is the potentiated startle response, produced by preceding the startle stimulus by either a secondary punishing (Brown 1961) or secondary frustrative (Wagner 1963) stimulus. Fear-potentiated startle is blocked by benzodiazepines, barbiturates, and buspirone (Chi 1965; Davis *et al*. 1993) when these are given during the test as opposed to acquisition phase.

In general, then, we can make the case that the anxiolytic drugs impair the functioning of the behavioural inhibition system as this was defined in the previous chapters (Fig. 3.1). The drugged animals care less about threats of punishment, omens of failure, or the uncertainties of a novel environment. In caring less, they inhibit behaviour less, they are less attentive to threatening stimuli, and they are less aroused. This description is surely one that applies to people who are being treated with anxiolytic drugs. Thus it is reasonable to conclude that *these drugs reduce anxiety in other animals as in man.*

Our description of anxiolytic drug effects has emphasized a change in mood or disposition. However, we have also seen data which suggest that, related to this alteration in mood, there is a cognitive factor. It is the cognitively complex tasks which demonstrate the clearest drug effects. Rather than discuss the issue of cognition in the present chapter, we leave it until we present the full theory (Chapter 10), by which time we shall have taken a detailed look at brain mechanisms of memory (Chapter 8). We also leave, until Chapter 11, discussion of the effects of the anxiolytics on memory. For the moment, we merely note the possibility that anxiolytic drugs produce indirect effects on the expression of anxiety by interfering with cognitive mechanisms.

The behavioural effects of the anxiolytic drugs, then, provide a reasonably coherent pattern which we can seek to see replicated by dysfunction of specific neural systems. It will be the business of later chapters to attempt to match these drug effects to some part or parts of the neurology of defence (see Appendices 2 and 3 for full details).

5 A theory of the behavioural inhibition system

This chapter outlines the picture of the behavioural inhibition system provided by the ethological, psychological, and pharmacological data so far, since it is this system that we take to be central to anxiety. In subsequent chapters we shall use this outline to guide our search for the neural basis of the behavioural inhibition system, and therefore of anxiety.

Let us first deal with the primary input to the system, the eliciting stimuli which activate it. In Chapter 2, we discussed the ethology of threat systems. Initially, following the Blanchards' analysis, we distinguished responses to immediate present threat from responses to a potential threat. In the paradigm cases studied by the Blanchards these two types of threat pose theoretically different functional requirements and, in practice, give rise to quite different types of behaviour. Active responses to a present threat are directed to avoidance of the source of threat. Active responses to a potential threat are directed to approach to (and assessment of) the source of potential threat.

A key aspect of the Blanchards' analysis is the idea of defensive distance. In the specific case of the potential threat paradigm, a small defensive distance (high expectation of danger) produces behavioural inhibition so great that there is no active behaviour; with a larger defensive distance, behavioural inhibition (i.e. inhibition of prepotent behaviour) is present, but is accompanied by active risk analysis and distinctive conflict-related behaviours such as the 'stretched-attend' posture; with very large defensive distance, both behavioural inhibition and active threat-related behaviour disappear and pre-threat behaviour returns. An important point was that active threat-related behaviour (risk assessment) showed an inverted $\cup$ relationship to defensive distance, and anxiolytic drugs had effects consistent with an increase in defensive distance. That is, they could increase or decrease active threat-related behaviour, depending on whether the defensive distance was smaller or larger, respectively.

Despite the apparent clarity of the Blanchards' analysis, there is a problem with viewing the categories of behaviour in terms solely of their nominal eliciting stimuli. Certainly, the response pattern to a potential cat includes risk analysis and behavioural inhibition (that is inhibition of pre-threat behaviours) as the rat enters the threatening situation; and both of these are sensitive to anxiolytic drugs. However, this response pattern also includes active escape from the threatening situation of the same form that occurs in response to an actually present predator. In a similar manner, the very real, as opposed to potential, presence of an immovable shock prod can elicit cautious approach and defensive burying, which is also sensitive to anxiolytic drugs. Furthermore, the animal's response to a discrete, non-moving, source of shock appears to depend markedly on how distinctive the source of shock is. A poorly localized source elicits similar behaviour to that which would normally be elicited by a potential predator.

Our resolution of these difficulties has been to move from a classical behaviourist stimulus-bound position to a more contemporary cognitive and functional view of the situation. We proposed that, on the one hand, we have a fight–flight–freezing system, which controls behaviour whenever the animal's primary purpose is to remove itself *from* a source of danger—that is, when its goal is to reach safety. On the other hand, we have a behavioural inhibition system, which controls behaviour whenever the animal's primary purpose is to achieve some goal which requires it to move *towards* a source of danger—that is, when it has *concurrent conflicting goals*: of reaching safety and of satisfying appetite. Thus the Blanchards' analysis is based on the distinction between an actual and a potential predator. In partial contrast, our view is that the predator/potential predator distinction at the stimulus level is highly correlated with, but not a necessary condition for, activation of the fight–flight and behavioural inhibition systems, respectively. Rather, the fundamental distinction depends on whether the animal's actions are aimed at entering (behavioural inhibition system) or leaving (fight–flight system) a dangerous situation—a matter of 'defensive direction'. As in the first edition, we identify anxiety with activity in the behavioural inhibition system; but this identification can now be seen as an extension of, and enriched by, the Blanchards' ethological analysis of responses to a potential predator.

The ethological perspective on the behavioural inhibition system has the advantage of functional clarity in its dependence on innate responses in ecologically reasonable circumstances. However, it is at a disadvantage if we want to gain greater insight into the cognitive processes that can give rise to the animal's decision to approach or to avoid a source of danger. In Chapter 3, therefore, we looked at a learning-theory analysis of the capacity of threatening stimuli to support new learning. As with the ethological analysis, but for different reasons, we found it necessary to categorically differentiate active avoidance from behavioural inhibition. The results of learning experiments also led us to expand the concept of the behavioural inhibition system from that of a system specifically activated by the requirement to approach danger to a system generally activated by any approach–avoidance conflict. We concluded that such conflicts could be produced not only by the innate fear stimuli analysed by the ethologists and not only by learned signals of impending punishment, but also by signals of impending frustrative nonreward and by novelty. Crucially, it is not the presence of the aversive stimuli themselves which activates the behavioural inhibition system, but their conjunction with appetitive stimuli or other conditions that result in the animal's having to choose between conflicting, incompatible goals. Novelty as a stimulus is special in this respect in that the same stimulus can be presumed often to give rise to both the approach and the avoidance tendencies constituting the conflict.

At this stage of the analysis, then, the behavioural inhibition system was viewed as a system which is engaged during approach to aversive stimuli. From this point of view we can derive both its normal adequate stimuli (signals of punishment, signals of frustration, novelty, and innate threatening stimuli—*all when they must be approached*) and its normal outputs (inhibition of prepotent behaviour, arousal, attention, and, where appropriate, risk assessment).

In both Chapters 2 and 3 we used the effects of anxiolytic drugs as additional evidence for the categorical distinctions we had made on ethological and learning-theory grounds.

These compounds appear to leave the functioning of the fight–flight system intact while quite generally impairing that of the behavioural inhibition system in a manner consistent with an increase in perceived defensive distance. Since the drugs concerned were defined by human clinical work, and the behavioural categories by two quite independent analyses of the behaviour of undrugged animals, this matching up of anxiolytic drug effects with anxiety-related behaviour is not tautological. (Indeed, in a few minor cases, like the anxiolytic reduction of vocalization in the presence of a predator, discussed in Appendix 1, the match is not perfect.)

Since Chapters 2 and 3 suggested that the common actions of the different classes of anxiolytic drugs could be used as a probe for anxiety itself, we reviewed in Chapter 4 the effects of the anxiolytic drugs on both clinical and animal behaviour (dealt with in more detail in Appendix 1). An important outcome of this review is the congruence between the effects of classical and novel anxiolytics, despite their different mechanisms of action and profiles of side-effects. The material reviewed in Chapter 4 substantiated the picture of anxiety involving, and anxiolytics reducing, behavioural inhibition, risk analysis, and a variety of other responses resulting from approach–avoidance conflict. However, it also showed that approach–approach and avoidance–avoidance conflicts which require behavioural inhibition (or produce it in practice) are also affected by anxiolytic drugs. This broadened concept of conflicting goals and behavioural inhibition included a number of effects in what would normally be considered tests of memory. Full analysis of these tests was postponed until Chapter 8; but the effects in them of anxiolytics suggest that the actions of these compounds in non-memory tasks may be exerted on the cognitive processes that ultimately lead to emotional responses, as much as directly on the emotional responses themselves.

Chapter 4 also substantiated the distinction between the fight–flight system and the behavioural inhibition system. The anxiolytic drugs are ineffective in reducing innate or conditioned escape or active avoidance responses. We were, in addition, able to distinguish the behavioural inhibition system from the behavioural approach system, since anxiolytics do not affect appetitive learning or a wide variety of other responses which are uncontaminated by the presence of conflicting goals. The only potential exceptions to this rule are a subset of memory tests (which, as argued in Chapter 8, must involve at least one of a number of types of goal conflict if they are to be sensitive to the drugs). There was also confusion surrounding the results obtained in several of the more recent screening tests for anxiolytic drugs. This could be because these tests are not firmly grounded in classical learning theory and, as a result, often contain elements which could activate both the fight–flight and behavioural inhibition systems concurrently. In the case of the novel anxiolytics, a further potential confound is the interaction between these compounds and the activity of the pituitary–adrenal system.

Putting Chapters 1–4 together, then, we have an updated view of the basic psychological components of the behavioural inhibition system (Fig. 5.1; cf. Fig. 1.1). This system is most characteristically involved in, and may have originally evolved to cope with, the requirement to enter threatening situations. However, potential predatory threat is not the only situation which engenders approach–avoidance conflict, and all approach–avoidance conflicts have many functional features in common. Either originally, or subsequent to its initial evolution, the behavioural inhibition system broadened its scope from

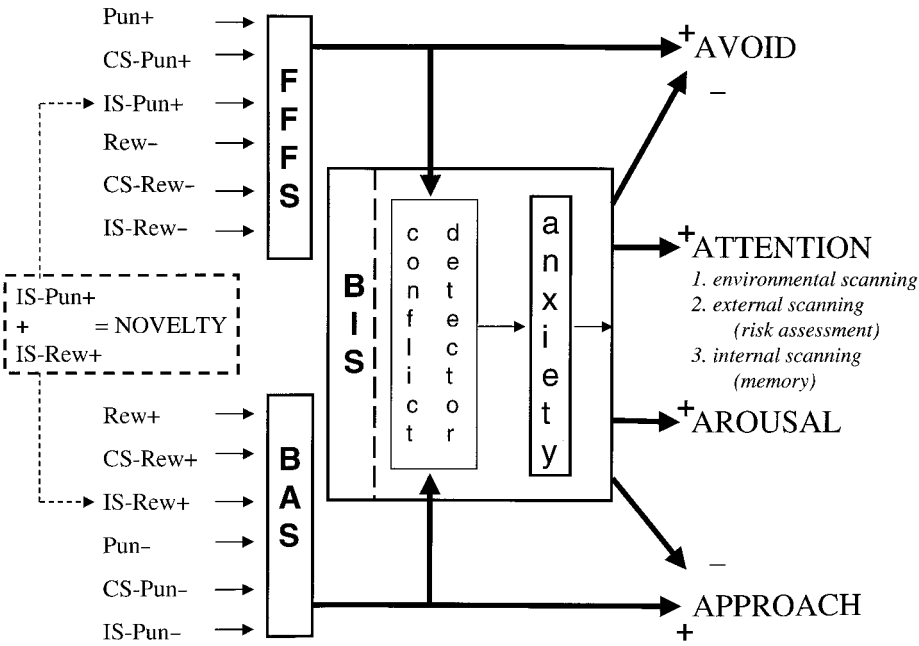


Fig. 5.1 An expanded view of the behavioural inhibition system (BIS) (cf. Figs 1.1 and 3.1; see Chapter 3 for definition of symbols) showing its relation to the fight–flight–freezing system (FFFS) and the behavioural approach system (BAS). As indicated by this diagram, the simplest way to activate the BIS is to concurrently activate the FFFS and the BAS, i.e. face the animal with an approach–avoidance conflict. In this case both simple approach and simple avoidance will be inhibited and replaced with environmental scanning (in the form of altered attention), external scanning (risk assessment behaviour), and internal scanning of memory. Note that all of these scanning operations are aimed at detecting affectively negative information and involve an increase in the salience of such information. As a result, a secondary consequence of activation of the system is a shift of the balance between approach and avoidance tendencies in the direction of avoidance.

potential predators to all the aversive stimuli which we can at present envisage as requiring resolution of approach–avoidance conflict: innate and acquired signals of punishment, innate and acquired signals of frustration, and initially threatening novel stimuli. In sum, the behavioural inhibition system functions to resolve conflicts between approximately equally activated and incompatible goals.

In Fig. 5.1 we have made explicit the requirements for approach–avoidance conflict for each of the different classes of stimuli that were to some extent implicit in Fig. 1.1 of the first edition. This change of emphasis is important, not only because it simplifies the description of the adequate stimuli for activation of the behavioural inhibition system, but also because it highlights the fact that the presence of stimuli or contingencies *per se* is not sufficient to activate this system. The animal's knowledge of those stimuli and contingencies must be such as to engender a genuine conflict between mutually incompatible goals (e.g. safety and food). In an example of the converse that we have used before, once an animal has learned to avoid shock, its behaviour is controlled

by habit and the behavioural inhibition system is no longer involved. Thus the operation of the behavioural inhibition system is fundamentally concerned with the 'hesitation and doubt' of Simonov's quotation with which we prefaced this book.

Not only the inputs but also the outputs of the behavioural inhibition system can be derived from the requirement to correctly resolve conflicting goals. Given an approximately equal balance between appetitive and defensive behaviour, neither is appropriate. Gluttony may turn you into someone else's lunch; timidity may starve you to death. The first requirement then is inhibition of both the behavioural approach and the fight–flight systems—the 'behavioural inhibition' from which we have named the behavioural inhibition system. This provides a preliminary resolution to the appetitive–aversive conflict by permitting neither of the incompatible goals to gain immediate control over behaviour. The second requirement is an increase in arousal: one must be prepared for split-second changes between approach to the appetitive goal and escape from a potential danger that may suddenly become manifest. The third requirement is that of increased attention. This may be relatively passive and involve increased scanning of the environment, or it may involve a wide range of risk assessment behaviours. To anticipate somewhat, it could also involve increased 'internal scanning', that is, increased retrieval of associations from memory coupled with their assessment for threatening or otherwise adverse implications.

5.1 EVOLUTION, ANXIETY, AND RULES OF THUMB—THE SEARCH FOR THE BEHAVIOURAL INHIBITION SYSTEM

Given this behaviour-based sketch of the behavioural inhibition system, our next step will be to try and find its neural substrate. In obtaining this sketch, we have already deployed the basic principles which will guide us also in the search for the neurology of anxiety: where both classical and novel anxiolytics affect behaviour in the same way, that behaviour is likely to involve elements of anxiety. Similarly, where both classical and novel anxiolytics produce a common change in neural function, that change is likely to involve the neural substrate of anxiety. In Chapter 6 we then use the effects of the anxiolytic drugs as a marker for the neural systems involved in anxiety.

To understand where this approach can lead, we need first to take a close look at the implications of the functional analysis just carried out. At the core of this analysis is the idea that the behavioural inhibition system is specifically adapted to the processing of one aspect of threat. Not only is threat phylogenetically common, but its physiological and behavioural requirements are similar across species (e.g. those required for successful escape). The species studied with anxiolytic drugs include fish, birds, mice, cats, dogs, pigs, and monkeys, as well as people. It is striking (Gray 1977) that, despite this diversity, there is virtually no need to qualify any of the resulting conclusions with respect to species. Where there is such a need, this is because of the intrusion of a species-specific complicating factor superimposed on the basic effect of interest, not because of an inherent lack of replication of the effect in that species. Furthermore, there is no sign of any change, qualitative or quantitative, in the effects of the anxiolytic drugs as one approaches, phylogenetically speaking, our own species.

This implies that the substrate upon which the anxiolytic drugs act includes components that are phylogenetically old; old enough to be present (so far as we know) in all contemporary mammalian species and to have homologues in all vertebrates. Birds and fish respond to anxiolytic drugs in ways which appear much the same as do rats and monkeys. This conclusion, based on behavioural experiments, is supported by studies of the benzodiazepine receptor. This appears to be present in the brains of vertebrates, including fish, quite generally (Nielsen *et al.* 1978). These data give us good reason in the specific case of anxiety, as do other data for emotions (LeDoux 1996) and mental functions generally (McNaughton 1989a, Chapter 11; Zinbarg and Mineka 1991; MacPhail 1996; Steckler and Muir 1996; Thomas 1996), to treat human beings as fundamentally similar to other mammalian species.

If the action of the anxiolytic drugs in so many diverse species is to reduce anxiety (and that is what the data suggest), it follows that anxiety itself is phylogenetically old. This greatly weakens any attempt to explain the *fundamental* mechanisms of human anxiety in terms that are specific to human beings (by recourse, say to the vagaries of the Oedipus complex), let alone specific to the pressures of modern life. Second, we now have an important clue as to at least some of the brain areas which must mediate human anxiety and on which the anxiolytic drugs act. These are likely to include, as major components, systems that are common at least to most mammals and likely, therefore, to lie outside the cerebral cortex which has achieved most prominence in ourselves and other primates.

Of course, this does not mean that phylogenetically recent structures do not play a major role in anxiety among the creatures that possess them. Even in the simpler organisms, conditioning and expectation play major roles in the production of anxiety. It follows that the range of stimuli which can elicit anxiety must depend on the perceptual and cognitive equipment available to the species in question. (To take a trivial case, the rat, being colour blind, cannot respond differentially to a red as opposed to a green stimulus that predicts a shock, while a pigeon can.) Nonetheless, given the evidence from studies of the anxiolytic drugs, it is virtually certain that the core mechanisms involved share much in common between human anxiety and the anxiety experienced by a rat or mouse, enough at any rate for it to be reasonable to search in the brains of these small creatures for the neural mechanisms that mediate human anxiety. In fact, in our final theory the phylogenetically oldest mechanisms are assigned a general modulatory function which is, if anything, more important for more highly encephalized species, while being chemically and functionally the same across all species. By studying these mechanisms in less encephalized species we can get a clearer picture of their function than we could in people (and indeed in other primates).

Perhaps because of their phylogenetic ubiquity, benzodiazepine receptors are remarkably widely distributed in the brain. Despite this, there is no evidence that the putative endogenous ligands of these receptors are released as neurotransmitters or even neuromodulators. An alternative possibility is that these ligands consist of a family of hormones or neurohormones able to affect multiple spatially separated targets in the brain (and possibly the rest of the body as well). Evolution often proceeds through phylogenetic conservation of simple mechanisms combined with elaboration of function by the progressive addition of further simple mechanisms; these are then able to extend the range of

circumstances to which the animal is adapted or otherwise to modify the operation of the existing mechanisms. A process of this kind might give rise to multiple sites of action of the anxiolytic drugs, sensitivity at each site arising from its own combination of random mutation and selective advantage. This possibility is consistent with our ethological analysis. This has shown that, even when we restrict ourselves to the behavioural inhibition system in particular as opposed to defensive reactions in general, we are dealing with a complex system with multiple effector components, each of which has been subject to somewhat different adaptive pressures (see Marks and Nesse 1994; McNaughton 1989b).

We have previously argued (McNaughton 1989a) that each emotion is best defined as a set of reactions to its own unique historical functional class of situation (see also LeDoux 1996). In the case of at least the 'basic' emotions (Ekman 1982; Ekman and Friesen 1986), there are two points to note. First, a large part of the form of the emotional reaction is the result of evolution (although this does not mean that it is rigidly fixed). Second, most emotions (and certainly anxiety) involve action tendencies, autonomic responses, hormonal responses, and probably immune responses. Thus, all the effector systems of the brain are likely to be involved simultaneously. However, it does not follow that this plethora of activity which we call 'an emotion' must be produced via the activation by a single environmental stimulus of a single central state which then triggers all the responses of all the effector systems. If that were the case, we would probably have had a generally agreed theory of the emotions a century ago. Rather, there are good evolutionary reasons for believing that most, if not all, emotions give only the appearance of a single source of central control. The pattern of effector output to which we assign a specific emotion label, such as 'anxiety', could result more from consistencies in the functional characteristics of the external eliciting stimuli than from the presence of some single locus in the animal's brain controlling all the outputs of all the co-activated effector systems.

Let us illustrate this point with the example used by William James (1884). You are faced with a lion (or any animal is faced with a predator). Your capacity to recognize that the lion is dangerous may be the result of learning and the precise form of your fear reaction may have been modified by experience, but to a large extent your 'motor tension, autonomic hyperactivity, apprehensive expectation, and vigilance and scanning' (DSM-IIIR, American Psychiatric Association 1987) are determined by the fact that your ancestors have frequently been faced with predators (if not lions) and that in the presence of *any predator* certain responses will increase the chances of survival. Your responses, therefore, are in part due to the prior meeting of random mutation with selective advantage. The necessary involvement of random mutation, here, has some interesting consequences for emotion. In particular, the response of each effector system will be subject to its own evolutionary equation; and, to the extent that the response of any effector system is genetically determined, it will be the result of a mutation which in many cases will be specific to that system. We have detailed the consequences of this state of affairs elsewhere (McNaughton 1989b), but they can be understood by brief inspection of Fig. 5.2.

Figure 5.2 shows, at the top, the logical situation to which we are referring. A range of stimuli S1–S9 can occur in the environment and, with the exception of S8 and S9,

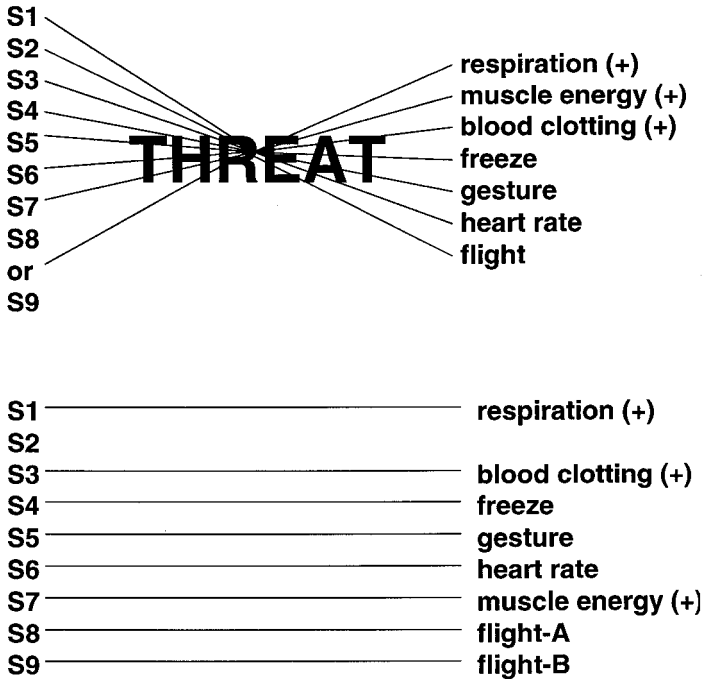


Fig. 5.2 The extremes of the possible neural relations which could have evolved to control responses to threat. The top half of the figure shows the functional relations linking stimuli (S1–S9) to responses where the stimuli may be regular predictors of threat (S1–S7) or where different stimuli are predictive of threat at different times (S8, S9). It can also be viewed as a representation of the simplest view of emotional states, namely that all stimuli activate a single neural representation of threat and this in turn activates the separate response systems. The bottom half of the figure shows, in its most extreme form, the opposite type of neural organization suggested by Hofer's experiments (see text). Here, each response system is under its own private stimulus control. (Redrawn from McNaughton 1989b.)

usually occur together. These could be, for example, the smell of the lion, the sight of its teeth, its outline, and so on. Each of these stimuli, alone, is a good predictor of threat, i.e. the presence of a predator or some other situation requiring the same general classes of response. Given that a threat *is* detected, each of the responses in Fig. 5.2 will, in general, be advantageous (though not necessarily for the same reasons; e.g. if escape is successful, there is no need for an increase in blood clotting factor). What is the probability that the neural (or mental) control situation will parallel the logical one represented by the top half of the figure? Is there a single central node which detects all forms of threat and triggers all effector systems? Given the normal workings of mutation and selection, an equally likely scenario will be that shown in the lower part of the figure, each effector system, and each response of each class of effector system, now being under separate control.

William of Occam would not be pleased with this suggestion, and evolutionary arguments are notoriously slippery, so what evidence do we have that such a complicated

view might be necessary? One example should suffice to show that we must at least take the possibility seriously.

Separation anxiety occurs in both children and the young of other mammals such as rats, dogs, and primates. When the 'reaction is beyond that expected for the child's developmental level', it becomes Separation Anxiety Disorder (DSM-IIIR, American Psychiatric Association 1987, p. 58). In the case of rats, separation anxiety is manifest in both behavioural and autonomic responses. These occur in concert when the mother is removed and appear, therefore, to be different effector outputs of a single, unified central state. Certainly, one could argue that, if either output were missing, the result would not be separation anxiety. However, it has been shown that the behavioural reactions can be eliminated by the presence of a non-lactating foster mother, while the autonomic reactions can be eliminated by feeding with milk, but not in either case vice versa (Hofer 1972). Thus, the two effector aspects of the 'one emotion' can be doubly dissociated in the laboratory. This does not make the label 'separation anxiety' any less useful, but it does mean that we must view the term as referring only to a particular class of evolutionarily recurring situation (loss of parents) which gives rise to a regular effector pattern (behavioural and autonomic) and hence a consistent central state, but not as referring to, or in any way implying, a single central control mechanism. The source of the apparent unity in ecologically normal circumstances is that removal of the mother necessarily eliminates simultaneously both her behavioural and her nutritional support of the pups.

Thus, both for the individual pattern of reactions making up a 'single emotion' on a single occasion and for the patterns making up 'the same emotion' on separate occasions, we must be prepared for apparent coherent, integrated activity to result from the activation of a number of what have been termed 'rules of thumb' (see Krebs *et al.* 1983). The tension between this type of relatively fragmented solution to the problem posed by the neuropsychology of anxiety and the more unified position adopted in the first edition of this book will occupy our attention at several later points in the development of our argument (see especially Chapter 6, Section 6.3.6, and Chapter 11, Section 11.3, and the final two chapters).

5.2 BEHAVIOURAL INHIBITION AND RULES OF THUMB

A 'rule of thumb' is a simple rule which delivers approximately 'optimal' behaviour, usually within a limited stimulus domain. In some cases (even with single-celled organisms), the output from the simple rule may be indistinguishable from theoretically optimal behaviour until the organism is placed in an artificial environment. In many cases, however, any one rule of thumb will only deliver adaptive behaviour in a limited domain. But, given sufficient mutations, a number of such rules can accumulate, each working within a limited domain but selected for because it extends the environmental range of the adaptive behaviour. In the end, a sufficiently large set of rules of thumb will give the appearance of a single coherent adaptive strategy applied over the whole of the animal's usual ecological range.

We have already considered the case where two rules (one for behaviour and one for the autonomic system) can be used concurrently in the same situation. Behavioural

inhibition supplies us with an example of two rules (both for the behavioural system) which give apparently similar output in different situations.

As discussed in Chapter 4, anxiolytic drugs decrease behavioural inhibition of lever-pressing on schedules of successive discrimination, fixed interval and differential reinforcement of low rates of response (DRL). In each case the behavioural inhibition can be characterized as resulting from a conflict between a prepotent lever-press response for food and a signal of reward omission (the negative sensory stimulus in successive discrimination, the delivery of reward in the fixed interval, the previous lever press in DRL). At the level of learning theory there is no reason to distinguish the three cases. In each case, furthermore, anxiolytic drugs of all types (including buspirone with appropriate doses and time courses of administration) decrease behavioural inhibition by increasing responding suppressed by the signal of non-reward more than they increase responding in the rewarded portion of the schedule. Thus, at the pharmacological level also, there appears, at this point, no reason to distinguish the three cases. However, a surprise awaits if we challenge the effects of anxiolytics in these tasks with the opiate receptor antagonist naloxone. Naloxone blocks the effects of the anxiolytic benzodiazepine chlordiazepoxide on the DRL schedule, does not alter the effect of chlordiazepoxide on successive discrimination (Tripp and McNaughton 1987; Tripp *et al.* 1987), and blocks the effect of chlordiazepoxide in the early part of the fixed interval, but not the later part (Fig. 5.3). We can conclude from this pattern of results that behavioural inhibition (as a specific output of the behavioural inhibition system) is mediated by at least two distinct neural pathways, of which one is sensitive to anxiolytics but does not contain an endogenous opiate link, while the other is sensitive to anxiolytics and does contain such a link. In some tasks (exemplified here by fixed interval responding) an apparently constant amount of behavioural inhibition (sensitive to anxiolytic drugs) can be under varying amounts of control by these different systems, interleaved so neatly that we cannot without special pharmacological manipulation see any break where one takes over from the other.

In our search for the neuropsychology of anxiety, then, we must be prepared for the possibility that we are dealing with a set of parallel systems, or a network of nodes, with distributed control, rather than a system with a linear flow of information through a single neural coordinating centre. We must be prepared also for the possibility that the surprisingly specific effects of the anxiolytic drugs on what we have identified behaviourally as anxiety is, paradoxically, the result of their capacity to affect simultaneously a number of quite separate nodes within such a parallel distributed system. Thus, while our psychological characterization of the behavioural inhibition system as a functional entity appears to be coherent and successful, this system may nonetheless consist of a number of differentiable components which work in parallel, although most likely with coordinating links both within the animal and via the environment. In short, there need be no central coordinating node. Indeed, the success of parallel connectionist models in the areas of perception and cognition make it very likely that control will be distributed across a neural network made up of multiple interacting nodes, no one of which is dominant.

The potential complexity of the neural systems involved in mediating anxiety does not mean, however, that a search for them must be for the proverbial needle in a haystack. On the contrary, our search will be highly focused by the actions of the anxiolytic drugs.

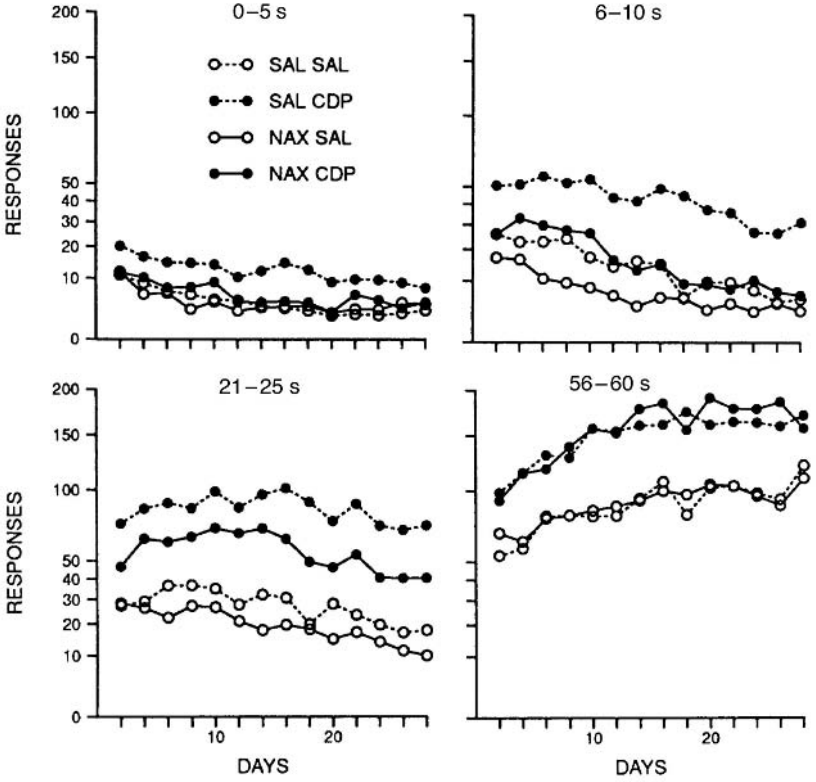


Fig. 5.3 Interaction of the opiate antagonist naloxone (NAX) with the anxiolytic benzodiazepine chlordiazepoxide (CDP) on a fixed interval schedule. CDP produces an increase in responding in all parts of the fixed interval. NAX blocks the increase totally at 0–5 s post-reward, not at all at 56–60 s after reward, and has intermediate effects at intermediate intervals. For further discussion, see text. SAL = saline. (From Tripp and McNaughton 1992.)

Specifically, we shall look for neural systems, activity in which is altered by the anxiolytic drugs and dysfunction in which produces behavioural effects similar to those of these compounds. Nor need we search at random for structures fulfilling these criteria. Given the phylogenetic arguments rehearsed above, the logical place to start our search will be in the limbic system. This is complex enough to subserve the functions we have delineated and phylogenetically old enough to be a realistic common substrate across many organisms. However, in our ethological, learning-theory, and pharmacological analyses we have also found it useful to link the operations of the behavioural inhibition system with those of the fight–flight system, acting, as they do, to control different modes of response to similar classes of situation. In the next chapter, therefore, we approach the deeper aspects of the limbic system from the direction of the periaqueductal grey, hypothalamus, and amygdala, subcortical structures known to be involved in the integration of primary defensive reactions.

6 The neurology of anxiety

The analysis pursued in Chapters 2 and 3 distinguished two classes of defence resulting from what could be called 'fear' and 'anxiety' respectively (although, as we saw, both these terms can be used to mean many other things). The simplest class of defence requires only escape or aggressive behaviour controlled by the fight–flight–freeze system. The second class is more complicated, since simple escape or avoidance tendencies are in conflict with a tendency to approach the source of potential threat. This type of approach–avoidance conflict is controlled by the behavioural inhibition system.

Where there is approach–avoidance conflict, the output of the fight–flight system (as well as appetitive responses controlled by the behavioural approach system; Gray *et al*. 1991) must be restrained so that defence is based on active risk analysis controlled by the behavioural inhibition system. This implies that the behavioural inhibition system provides an extra, inhibitory, level superimposed upon the fight–flight system. It seems likely, therefore, that the neurology of fear will be bound up in the neurology of anxiety. A further complication of anxiety is that the avoidance component of the approach–avoidance conflict need not involve fear, since we concluded that anxiety can also be based upon the risk of frustrative non-reward. However, fear is the most experimentally tractable and best understood precursor of anxiety, and so we shall analyse the neurology of this, the primary defence system, first before going on to the neurology of anxiety. Luckily, the neurology of the primary defence system has been worked out in considerable detail in recent years. It turns out, moreover, that the higher levels of this system control not only fear but also (unconditioned) frustration.

The tight linkage between fear and anxiety, as well as the potential complexity of fear-related systems, is underscored by the fact that simple classical conditioning of fear to contextual cues produces noticeable reactions in much of the allocortex, isocortex, subcortical telencephalon, thalamus, hypothalamus, and many areas of the brain stem; and, in the majority of these areas, peripheral administration of the anxiolytic benzodiazepine, diazepam, reduces (but does not eliminate) the reaction (Beck and Fibiger 1995). (Interestingly, the central nucleus of the amygdala, which many would regard as the heart of anxiety control, is the only place in which these authors observed an *increase* in activity with administration of diazepam.) As a result, detailed analysis of even the primary defence system would require a book at least as large as the present one. We resolve this problem in two ways: (a) in the present chapter, we provide only a general argument, leaving more detailed review to Appendix 2; and (b) even in Appendix 2 we do not attempt to provide a comprehensive overview, for two reasons. First, in the key area of the amygdala there have recently appeared a number of excellent reviews (Davis 1992a,b,c; Aggleton 1993; LeDoux 1994, 1995), as well as a volume edited by Aggleton (1992; and see also LeDoux 1996, for an amygdalo-centric view of emotions in general) on which we can and shall draw heavily. Second, while there are controversies about precise details, there is broad agreement about the general organization of the primary

defence system. We feel, therefore, that we can present a brief summary of what is, largely, a consensus view of the primary defence system and refer to the relevant reviews for the extensive analysis on which it is based.

The picture we present is of a system hierarchically organized in terms of both motor control and stimulus analysis; and the structures on which we focus represent the interface between stimulus and response, so that what their neurons code for is probably best seen in terms of goals.

The hierarchical organization we describe for motor control (Fig. 1.8; Table 6.1) is based on that proposed by Graeff (1994; see especially his Table 1), but with some minor changes and additions noted below. Briefly, we take as a starting point the hierarchical organization of levels of defence developed by the Blanchards (Chapter 2) and used as a foundation by both Graeff and ourselves. Fight, flight, and freezing are at the lowest level and result from a contacting or proximal danger, circumstances in which there is little time for analysis of the situation or freedom of action. Directed escape is at the next level and results from a more distal but clearly present danger. Here there is more time for analysis of the situation, a greater variety of possible responses and outcomes, and some room for conditioning. Next, there is active avoidance resulting from a potential danger which need not be approached. (This is a category omitted by Graeff.) Here analysis and motor programming must reach high levels of abstraction and will often involve conditioning. Next (above and beyond what is organized by the primary defence system), there is behavioural inhibition (conflated with freezing by Graeff) plus risk assessment resulting from a potential predator which must be approached, with the result that a high level of conflict is generated. This set of contingencies requires not only processing of information by both the structures that mediate avoidance and those which mediate the competing approach, but also additional highly complex information processing and priming of action tendencies so as to resolve the discrepancy between the outputs of the approach and avoidance systems: in short, it requires activation of the behavioural inhibition system.

Following Graeff (1994) we see this logical hierarchy at the psychological level as mapping almost directly onto the largely hierarchical anatomical and functional organization of the neural systems involved. This is shown in Fig. 1.8. There are two minor points where Fig. 1.8 diverges from the specific details of Graeff's proposal.

Our first divergence is that, as discussed earlier, we do not equate freezing with behavioural inhibition, and so we do not see the septo-hippocampal system as mediating freezing responses. The highest intensity of freezing occurs in response to a proximal predator, and usually precedes explosive attack or flight. Like Graeff, we believe that both high-intensity freezing and explosive attack are mediated by the periaqueductal grey. This freezing response forms an integral part of a set of behaviour patterns which remove the animal from the dangerous situation. As noted earlier (Chapter 2), the problem with any distinction between high-intensity behavioural inhibition and at least low-intensity freezing is that both are most easily characterized by an absence of behaviour. This, no doubt, is the basis for Graeff's amalgamation of the two. Our own analysis bears a greater burden of theory. We use the term 'behavioural inhibition' to indicate that there is indeed a form of behaviour that is being inhibited; in particular, a form of approach behaviour. In the theoretically simplest case, this is approach to an appetitive

goal (e.g. food, water, or merely a part of the environment that might contain these valuable commodities). In a more complex case, the approach behaviour that is inhibited is approach to safety, as in a learned escape or active avoidance response, which (like approach to an appetitive goal) we see as mediated by the behavioural approach system. Thus, in our usage, 'behavioural inhibition' does not necessarily imply a cessation of all behaviour. Indeed, it can, and often does, occur concurrently with *active* risk assessment. It is only at the shorter defensive distances that complete behavioural inhibition occurs. As we have seen, behavioural inhibition, in this sense, appears distinguishable from freezing by being sensitive to anxiolytic drugs (Chapter 2) and, less tautologically, by the body postures adopted, e.g. the 'stretched attend' posture which would not be seen in true freezing. Blanchard and Blanchard (1996) have suggested the term 'defensive quiescence' to refer to this type of behaviour.

We can, then, distinguish two types of stillness, both of which may be referred to in the literature as 'freezing'. The first could occur, for example, when the animal is awaiting the optimal moment to emit an innate flight response. Here there would be freezing, and so superficially behavioural inhibition, but there would be no tendency to approach available appetitive goals and the stillness would reflect a high level of activity in the fight–flight system but not in the behavioural inhibition system. The second could occur, for example, when the animal is awaiting the optimal moment to approach a dangerous goal. Here, there would be stillness, as before, but there would, in addition, be a definite preparedness to approach an appetitive goal, and so the stillness would reflect a high level of activity in the behavioural inhibition system and an accompanying inhibition of the outputs of both the behavioural approach and fight–flight systems. A reliable experimental means of making this distinction would be of great value. Provided, however, that it is real, we may retain the notion of defensive direction (i.e. the postulate that the involvement of the behavioural inhibition system and of the emotion of anxiety depends upon the need to approach danger) which characterized our analysis in Chapters 2 and 3. As a bonus, this analysis also allows a clean separation of levels in terms of neural structures (compare Graeff's, 1994, Table 1 with our Table 6.1) distinguishing the different psychological levels of the defence system. Of course, such functional neatness is not a necessary property of brain organization, and further research will be required to settle the matter.

Our second divergence from Graeff's scheme is an implied branching in the hierarchical system. The septo-hippocampal system is neurally (see below) and logically 'above' the amygdala. However, as far as active avoidance tendencies are concerned, the anterior cingulate cortex is also 'above' the amygdala. In terms of a purely linear escape-based system, then, the strict progression is dorsal periaqueductal grey–medial hypothalamus–amygdala–anterior cingulate (distinguished by increasing active defensive distance), with septo-hippocampal system–posterior cingulate (in which the levels are distinguished by increasing passive defensive distance) constituting a side-branch off the main stem at the level of the amygdala (the two branches being distinguished by defensive direction; Fig. 1.8).

The separate, but also hierarchical, organization we describe for stimulus processing is largely that proposed by LeDoux (e.g. 1992, 1994, 1995, 1996). On this view, several levels of sensory processing are connected, independently, to the amygdala. For example,

Table 6.1 Levels of threat processing and related levels of neural integration (largely drawn from Graeff 1994; see text for modifications). (See also Fig. 1.8)

Level of processing	Behaviour	Neural substrate
Potential danger (to approach)	Risk assessment and behavioural inhibition	Posterior cingulate Septo-hippocampal system
Potential danger (to avoid)	Avoidance	Anterior cingulate Amygdala
Distal danger	Escape Inhibition of aggression	Medial hypothalamus
Proximal danger	Freezing Flight Fight	Periaqueductal grey

the amygdala receives input from both the visual thalamus and from the highest levels of visual cortex. The central underlying principle of this type of organization is illustrated in Fig. 6.1. The lowest levels of stimulus processing provide the amygdala with only partially digested information, but they do so with greater speed than the higher levels. Having parallel inputs to the amygdala results in a number of useful properties. It allows extremely fast processing of the simplest and most distinguishable danger stimuli. It allows, in the case of complex or poorly distinguishable stimuli, for priming (rather than full-blown activation) of defence circuits by lower levels of processing to speed up the subsequent response to input from higher levels of processing. Finally, it allows for receipt by the defence system of information about danger from the very highest levels of cognitive analysis, even if such danger has not been detected by any of the lower levels.

> Animals, and humans, need a quick-and-dirty reaction mechanism. . . . Failing to respond to danger is more costly than responding inappropriately to a benign stimulus. . . . Details are irrelevant and, in fact, detrimental to an efficient, speedy and potentially lifesaving reaction. The brain simply needs to be able to store primitive cues and detect them. Later, coordination of this basic information with the cortex permits verification . . . or brings the reaction to a stop. (LeDoux 1994, p. 38.)

As we describe below, this parallel organization of perceptual inputs, coupled with the capacity for plasticity of synapses in the amygdala, provides a simple basis for the learned association of a wide variety of stimuli with defence systems. It also provides a basis for the evolution of innate threat stimuli.

An interesting point, when we compare the schemes of Graeff and LeDoux, is that the septo-hippocampal system figures in both. Graeff sees it as the top level of the hierarchical defence control system. LeDoux sees it as one of the highest levels of the threat analysis system, essentially a stage beyond polysensory neocortex. The picture we shall paint of the septo-hippocampal system to some extent combines both these points of

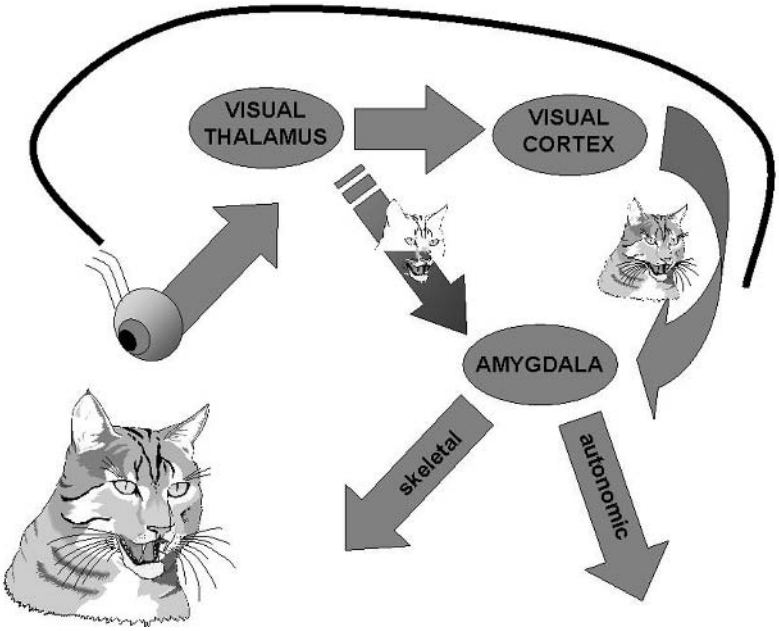


Fig. 6.1 'Quick and dirty' versus 'slow and sophisticated' processing of threat stimuli. The cat in the external world creates impulses at the retina which proceed to the visual areas of the thalamus and then quickly via a direct pathway to the amygdala. The representation in the thalamus will be very sketchy but will allow either priming, or generation, of escape responses with the minimum of delay. Slower processing, via the various levels of visual cortex will ultimately produce a more detailed and exact classification of the stimulus, which will then also be sent to the amygdala—allowing either completion of escape or, if the quick and dirty system produced a false alarm, cancellation of the escape response. (See LeDoux 1994, Fig. 2, for the snake/human original on which this cat/rat figure is based.)

view but, as noted above, places the hippocampus outside the primary chain of command. As discussed in Appendix 3, if we had to nominate a 'top' of the active defence system it would be the anterior cingulate cortex.

Having briefly described the primary defence system as a whole, we now provide a little more detail on each of its elements, starting with the lowest and proceeding to the highest. (These are all dealt with at greater length later: the dorsal periaqueductal grey, medial hypothalamus, and amygdala in Appendix 2; the septo-hippocampal system in Chapters 7–10 and Appendices 4–9; and the cingulate and prefrontal cortices in Appendix 3.) We shall be able to take much of the neurology of the lower levels of the defence system as given.

As noted in the previous chapter, the key tool we shall use to draw the (somewhat fuzzy) line between primary defence systems and systems mainly involved in anxiety lies in the behavioural effects of the anti-anxiety drugs (Chapter 4 and Appendix 1). We shall compare the resulting 'behavioural profile' of these compounds with that of specific lesions of each of the candidate structures (Table 4.2), in the belief that anxiolytic drugs are likely to act like a lesion of those neural structures most involved in the control of

anxiety. Where data are available, we also take particular note of the effects produced by injecting anxiolytic drugs directly into the structures of interest. (Detailed consideration of the advantages and problems of this strategy of drug–lesion comparison is provided in Appendix 8.)

6.1 THE PERIAQUEDUCTAL GREY, FIGHT, FLIGHT, AND FREEZING

The periaqueductal grey appears to be the lowest level of the defence system from which a range of environmentally dependent integrated defence responses can be elicited, and through which the control of such responses by higher level systems is mediated. It is not a homogeneous area but, with high levels of activation and a suitable object to which to direct responses, 'all of the behavioural signs of defence, including directed attack, are found within the periaqueductal grey' (Bandler and Carrive 1988). In particular, activation of this region can produce the explosive, undirected escape typical of an animal in close proximity to a predator. There is thus good evidence that the periaqueductal grey is a major component of the fight–flight system; that it is likely to be the lowest level of the fight–flight system before control is passed to nuclei concerned only with individual components of the defence reaction; and that, by itself, it does not control particularly sophisticated defence responses.

Graeff (e.g. 1994) views the periaqueductal grey as being the principal locus for generating panic in human beings. Consistent with both this view and the view that it is a key part of the fight–flight system, interference with the periaqueductal grey produces little of the behavioural profile of the anxiolytics. Furthermore, with the exception of conditioned hypoalgesia (tested only with benzodiazepines), injection of anxiolytics into this region has little effect. Since our interest is in locating possible neural sites of action of the anxiolytics, we shall not consider the periaqueductal grey further at this point. However, a more detailed discussion can be found in Appendix 2.

6.2 THE MEDIAL HYPOTHALAMUS AND ESCAPE

The medial hypothalamus is reciprocally connected to the periaqueductal grey and in many ways seems to share its functions. Indeed, matching Graeff's suggestion for the periaqueductal grey, Shekar (1994) has suggested that activation of the dorso-medial hypothalamic 'cardiostimulatory' area can give rise to panic. These suggestions are not mutually incompatible, as excessive activation of the medial hypothalamus could spill over into activation of the periaqueductal grey. (Later, we shall suggest a similar model for the generation of panic by even higher levels of the system.)

The medial hypothalamus differs from the periaqueductal grey, however, in that escape responses elicited from it lack upright jumping and are interspersed with rearing, suggesting a greater defensive distance and possibly some shift in defensive direction. Nonetheless, escape elicited from the medial hypothalamus is blocked by lesions of the periaqueductal grey (and not vice versa). The medial hypothalamus differs from the

periaqueductal grey also in that lesions here increase aggression, whereas lesions of the latter region decrease it. These and a range of other data (Appendix 2) suggest that the medial hypothalamus represents the next level up in a hierarchical defence system. The opposite effects of lesions of the two areas on, for example, aggression are then seen as resulting from the need for the hypothalamus to inhibit the lower (quicker, dirtier) periaqueductal grey when controlling behaviour.

While activation of the medial hypothalamus can increase experimental anxiety in various animal tests, it also increases escape and avoidance. Thus, the increases in, for example, rearing produced by activation are likely to be due to a simple increase in fear under conditions where this can conflict with prepotent responses. Such activation may also secondarily spread to higher levels of the brain, including not only the amygdala but also the piriform and entorhinal cortices.

There is some indication (as with the dorsal periaqueductal grey) of minor effects of anxiolytic drugs on the medial hypothalamus, but no reason to suppose that this is a primary site of their action.

6.3 THE AMYGDALA

We have reached the amygdala by following a descending defence control system back up into the brain. However, even more so than the hypothalamus, the *efferent* connections of the amygdala ascend to link it intimately with a wide variety of limbic and posterior cortical areas, and do not merely descend to link it with the medial hypothalamus, periaqueductal grey, and related areas. Whether or not these ascending connections are present to cope with defensive situations, it is clear that they allow the amygdala to influence stimulus processing, and hence potentially memory, quite extensively.

Lesions of the amygdala produce much wider emotional deficits (the Klüver–Bucy syndrome) than would be expected if it were a purely defensive structure (see, for example, LeDoux 1992). Thus, while the amygdala can certainly be viewed as the primary higher structure controlling responses to actual (as opposed to potential) threat (LeDoux 1994), it can also be viewed as the means whereby, quite generally, 'sensory stimuli are endowed with emotional and motivational significance. . . . Thus, when the amygdala is damaged, monkeys act tame in the presence of humans because the sight of a human is no longer coded as a threatening stimulus. Similarly, *dietary and sexual preferences change since environmental stimuli no longer elicit their normal affective responses*' (LeDoux 1992, p. 340, our emphasis). Also, the 'amygdala . . . influences maternal behaviour by relaying olfactory input to the medial preoptic area' (Numan 1994, p. 19) and appears to have a related transducer role in mating (Malsbury and McKay 1994; Minerbo *et al.* 1994; Kondo and Arai 1995) and paternal behaviour (e.g. Kirkpatrick *et al.* 1994b), as well, probably, as in salt control (e.g. Seeley *et al.* 1993), food intake (e.g. King *et al.* 1993; Crovetti *et al.* 1995) and social interaction (e.g. Adolphs *et al.* 1994; Borlongan and Watanabe 1994; Kirkpatrick *et al.* 1994a). In human beings it has been suggested that 'the amygdala plays important roles in memory and in the modulation of social and emotional behaviour' (Tranel and Hyman 1990, p. 349).

Some care must be taken, however, if we are not to ascribe too broad a set of functions to the amygdala. It is located within the temporal lobe and closely connected to areas such as the hippocampus and entorhinal and cingulate cortices. As a result, lesions to, or stimulation of, the amygdala will not only influence these other areas because of its connections with them, but can also affect fibres of passage within and at its borders. These fibres of passage can often be crucial for temporal lobe functions which are completely independent of the amygdala. The recent tendency to use injections into, as opposed to electrolytic lesions of, the amygdala is, therefore, to be welcomed. To note just one example: taste aversion conditioning is impaired by conventional, but not cytotoxic, lesions of the amygdala (see LeDoux 1992, p. 341), unless the cytotoxic lesions are made in the basolateral nucleus (Yamamoto *et al.* 1994).

In dealing with the amygdala, then, we shall follow a strategy which we apply later on a much larger scale to the septo-hippocampal system. We attempt to combine the anatomical, lesion, and physiological data into a coherent picture, in the hope that each discipline will correct the faults inherent in the others. We concentrate on the role of the amygdala in anxiety and omit, for example, any discussion of its possible role in the symptomatology of depression (e.g. Schulkin 1994; Schulkin *et al.* 1994). We start with anatomy because its results (as opposed to their interpretation) are least likely to be overturned by findings in the other disciplines.

6.3.1 **Anatomy of the amygdala**

The anatomy of the amygdala is reviewed in some detail in Appendix 2 (which follows closely Amaral *et al.* 1992). While it is an extremely complicated area, one can take a relatively straightforward view of its overall organization. Simplifying somewhat, this consists of the lateral nucleus, the basal nucleus, and the corticomedial complex, with information flowing through these structures, essentially unidirectionally, from lateral to basal to corticomedial (Fig. 6.2). The amygdala appears to receive, in the lateral nucleus, information which can be used to identify threat (and related affective situations). This information is sent directly, i.e. without influence from additional information received by the basal and central amygdala, from the lateral nucleus to the hippocampal formation (particularly the entorhinal cortex) and to the dorsomedial thalamus, as well as to the basal nucleus of the amygdala. In each case there is little feedback.

The basal nucleus has an unusually reciprocal (and unusually restricted) connection with the hippocampal formation and also has extensive reciprocal connections with the frontal and cingulate cortices. In the latter case, we suggest that the projection to the cortex provides information about affect, with the return projection providing information about current plans (essentially read-out from working memory). This reciprocity between the basal amygdala and the neocortex could allow a recursive updating of current plans in the context of the predicted affective consequences. When we discuss the septo-hippocampal system in detail, we shall argue for a roughly similar, but more hippocampopetal, organization. The basal nucleus also provides output to all levels of visual processing in the cortex. We suggest that this could provide an affective bias to the processing of incoming information. The main clearly unidirectional output of the basal nucleus is to the striatum; this could provide affective bias to ongoing motor programs and their modification.

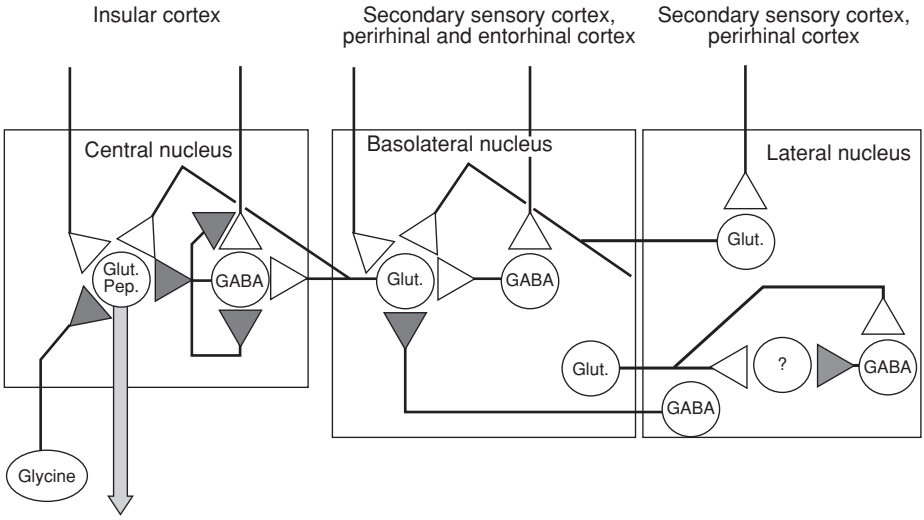


Fig. 6.2 Some of the internal connections of the amygdala. Information tends to flow from lateral to basal to central with each of these areas receiving distinct cortical connections. Glut., glutamate; Pep., peptide; GABA, γ-aminobutyric acid. Open triangles, excitatory synapses; filled triangles, inhibitory synapses. (From Davis *et al.* 1994; see also Fig. 6 in Pitkänen *et al.* 1997.)

The corticomedial complex receives information from both the lateral and basal nuclei. This organization raises the possibility that the corticomedial complex takes into account not only the primary affective information being determined in the lateral nucleus, but also the interaction between that information and concurrent processing in the frontal cortex, cingulate cortex, and hippocampal formation relayed via the basal nucleus. The primary output of the corticomedial complex, in contrast to that of the basal nucleus, is subcortical and appears to involve interaction with the 'lower' levels of affective control for which the amygdala represents one of the higher levels (see Fig. 6.3).

It has recently become evident that 'each portion of the prefrontal cortex has a distinctive projection to the amygdala. The ventral areas of the lateral and medial prefrontal cortices, which receive olfactory projections, are the only prefrontal cortical areas with projections to the olfactory-related superficial amygdaloid nuclei. The more dorsally situated prefrontal areas, the dorsal agranular insular area and prelimbic cortex, have complementary projections to the basal nucleus' (McDonald *et al.* 1996). The amygdala, then, may share the organization into a dorsal and ventral trend which we shall discuss in more detail for the frontal cortex (see below; and also Appendix 3).

6.3.2 **Effects of lesions**

The key question addressed here is how far the amygdala, which is clearly important for fear, is also the central structure for the elaboration of anxiety.

The effects of amygdalar lesions are reviewed in Appendix 2 and summarized in Table 4.2, where they are compared with the effects of anxiolytic drugs (considered above) and

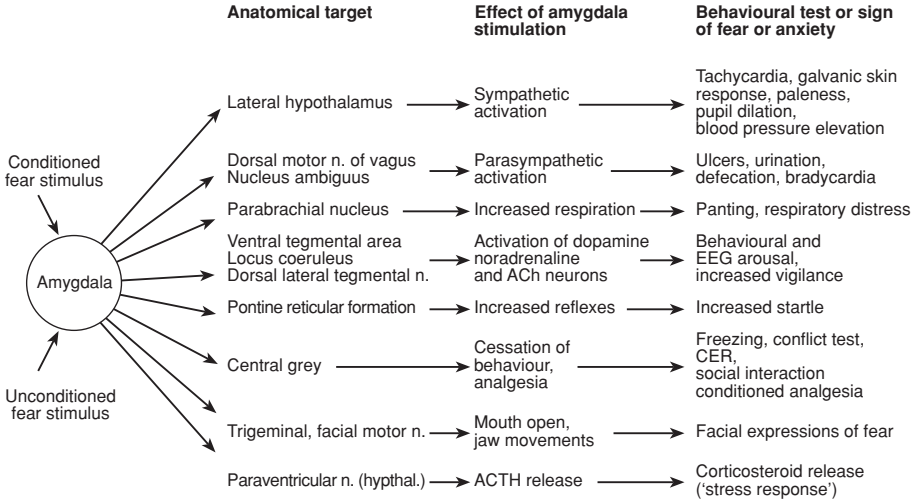


Fig. 6.3 Outputs from the amygdala. The amygdala can be viewed as the final stage where 'fear' is encoded before signals are sent to areas that deal with specific desired outputs. Thus amygdala lesion blocks all of the outputs listed while lesions of the anatomical targets of the amygdala produce only the indicated subsets of effects. n, nucleus; hypothal., hypothalamus; ACh, acetylcholine; ACTH, adrenocorticotrophic hormone; EEG, electroencephalogram; CER, conditioned emotional response. (From Davis *et al.* 1994.)

with those of septal and hippocampal lesions (considered shortly). The amygdala column of the table is critical for the logic of our argument. We accept that the amygdala is a crucial structure for fear. We shall also see shortly that anxiolytic drugs have some direct actions on the amygdala. Further, we argue later in this chapter that anxiolytics drugs could act indirectly on the septo-hippocampal system as a result of direct actions on the raphe and locus coeruleus. But this argument could apply equally well to an indirect action on the amygdala, which receives both serotonergic and noradrenergic innervation. Before we conclude, therefore, that the septo-hippocampal system is crucial for anxiety, we have to consider the possibility that it is the amygdala that plays this role, and that the effects of septo-hippocampal lesions (in particular their similarity to the effects of anxiolytic drugs) are a secondary consequence of hippocampal input to the amygdala.

As can be seen from Table 4.2, however, the list of behavioural effects of amygdalar lesions is both positively and negatively discrepant from the anxiolytic drug syndrome. Lesions of the amygdala decrease one-way active avoidance, decrease two-way active avoidance, increase rearing, and decrease the double-runway frustration effect. In all of these cases anxiolytic drugs (and, as discussed further below, septal and hippocampal lesions) are without effect or have the opposite effect. Equally, lesions of the amygdala are without effect on most delayed matching to sample tasks and on behaviour reinforced on fixed interval schedules. Both of these are affected by anxiolytic drugs (and, again, by septo-hippocampal lesions). It follows that, if anxiolytic drug effects are a good (inverse) guide to the nature of anxiety, anxiety does not easily map onto enhanced amygdalar function.

6.3.3 **Long-term potentiation and memory**

Long-term potentiation (see also Appendix 5) is a long-lasting synaptic enhancement first described by Bliss and Lømo (1973), which has now been described in many areas of the brain and spinal cord. It depends for its occurrence on the activation of voltage-gated channels, most commonly involving the interaction of the excitatory neuro-transmitter, glutamate, with the *N*-methyl-D-aspartic acid (NMDA) receptor. Binding of glutamate to the receptor or depolarization of the postsynaptic cell is, by itself, insufficient to activate the ion channel. What is required is that the cell be depolarized past a particular point *and* that then glutamate should bind to the NMDA receptor. Given this combination of circumstances, the ion channel opens and allows calcium to enter the cell. The calcium then triggers a complex biochemical cascade which can induce changes in both the postsynaptic cell and the specifically activated presynaptic terminal. The net result is that long-term potentiation is both input-specific and associative in the way required by Hebb (1949) for the formation of cell assemblies. While long-term potentiation is usually produced in the laboratory by non-physiological levels and patterns of stimulation, careful analysis of single cells has shown that it will occur under normal physiological circumstances and should involve changes in only the activated synapses on only those cells which are depolarized at the appropriate point in time (Abraham 1988).

There has been much effort expended (Appendix 2) in tracing the sensory inputs which provide the CS for conditioning of fear and the output circuits providing the unconditioned and conditioned responses, as well as in analysing these input and output circuits under a variety of electrophysiological and behavioural conditions. It has been reasonably well established that long-term potentiation of input to the amygdala provides the basis for at least some fear conditioning (Rogan *et al.* 1997), and that even quite complex stimuli probably produce fear reactions by the strengthening of input to the amygdala. Thus, it is generally accepted that the amygdala is a key structure in the organization of both unconditioned and conditioned fear responses. In the conditioned case, then, we have the rather neat picture of a structure which receives input from sensory areas and can link these to defensive output through the simple step of long-term potentiation. In this sense, emotional (or at least defensive) memory would be located in the amygdala (LeDoux 1994, 1996). There is some question as to whether the amygdala is additionally involved in non-affective memory. However, current evidence suggests that it is not, and that its apparent role in the relevant tasks is the result of damage to fibres of passage (but see Cahill and McGaugh 1996). This is not to say, however, that the amygdala is where, for example, conditioned avoidance reactions are ultimately stored. We argued earlier that emotion is involved only in the initial stages of avoidance learning, while the aversive UCS is still being delivered. The amygdala appears to be critically involved only in those initial stages.

6.3.4 **Single-cell responses**

The complexity of cellular responding in the amygdala is considered in some detail in Appendix 2. Many single cells respond, each relatively selectively, to a wide range of stimuli: visual, auditory, objects, faces, novelty, reward, punishment. There appears to

be a progression from relatively high stimulus selectivity in the lateral amygdala to relatively low selectivity in the medial amygdala (with the basolateral nucleus intermediate). The extent of response often seems tied to what could be termed the 'importance' of the stimulus. In particular, changing the affective associations of a stimulus frequently changes the response of the amygdala to that stimulus. It is likely, however, that the stimulus processing on which the amygdalar cellular response depends is carried out in other parts of the brain, since amygdalar firing seems to precede autonomic changes rather than to follow the stimuli that can elicit those changes. Thus, the firing of amygdalar cells may relate more to the programming of autonomic output than to the detection of particular complex events.

6.3.5 **Effects of anxiolytic drugs**

Since we have characterized anxiety as resulting from a conflict between avoidance and approach tendencies, it follows that damage to the amygdala should reduce anxiety—if it does this—by removing the capacity to detect threat and hence the tendency to avoid, acting in this way like the administration of an anxiolytic drug. As we saw above, there are a number of important discrepancies between the effects of anxiolytic drugs and those of amygdalar lesions. Nonetheless, some effects of the anxiolytic drugs appear to be achieved via a direct action on the amygdala. 'The amygdala has a high density of benzodiazepine receptors. . . . Local infusion of benzodiazepines into the amygdala (for review see Davis 1992a) has anxiolytic effects in the operant conflict test, social interaction test, measures of conditioned freezing and hypoalgesia [and] the light–dark box test in mice. . . . However, benzodiazepines can still have anxiolytic effects in animals with lesions of the amygdala' (Davis *et al.* 1994, pp. 212–3). Indeed these latter effects can be greater than in the absence of the amygdala lesions (Yadin *et al.* 1991, cited by Davis 1992a). Intra-amygdaloid benzodiazepine injection has also been reported to impair step-down passive avoidance (Tomaz *et al.* 1993; Harris and Westbrook 1995), shock-probe avoidance (Pesold and Treit 1994), conditioned suppression of drinking (Shibata *et al.* 1989), measures of anxiety and memory in the elevated T-maze (Tomaz *et al.* 1993), and thigmotaxis in an open field (McNamara and Skelton 1993). But injection of benzodiazepines into the amygdala normally has no effect on water maze learning (McNamara and Skelton 1993), on behaviour in the elevated plus-maze, or on defensive burying (Treit *et al.* 1993; Pesold and Treit 1994). The lack of involvement of the amygdala in the latter two cases is also shown by the fact that amygdalar lesions neither affect behaviour by themselves nor interact with the effects of systemic benzodiazepines on these measures (Treit *et al.* 1993). Despite the effect of amygdalar lesions on conditioned suppression of drinking, they do not appear to interact with the effects of benzodiazepines on this test either (Kopchka *et al.* 1992). More detailed analysis of parts of the amygdala may be required, however, since it appears that effects in the plus-maze and in passive avoidance can be obtained with injections into the basolateral but not central amygdala, while the reverse appears to be the case with shock-probe avoidance (de Souza Silva and Tomaz 1995; Pesold and Treit 1995).

We consider the relationship between amygdalar lesions and anxiolytic action again below. However, we note here that such lesions affect fear responses even when these

are uncontaminated with conflicting approach tendencies, while anxiolytic drugs do not. Conversely, infusion of benzodiazepines into the amygdala impairs response suppression resulting from shock but not from reward omission (Hodges *et al.* 1987). Overall, anxiolytic drugs seem to produce only some of their effects via the amygdala and, conversely, only some amygdalar processes are affected by anxiolytic drugs.

Given the critical involvement of fear as a component of anxiety, and the recursive connections between the amygdala and the septo-hippocampal system, we can account for the above findings as follows. The amygdala is the key structure for the coding of fear, i.e. the presence in the environment of a stimulus associated with pain. Accepting that the septo-hippocampal system is the key structure for anxiety, one way of reducing those aspects of anxiety which are dependent on fear (as opposed to other aversive sources of conflict) is to decrease the fear signal sent by the amygdala to the hippocampus. One way of achieving this would be by action at benzodiazepine receptors in those parts of the amygdala which project to the hippocampal formation. Consistent with this argument is the fact that these receptors are dense in the basal and lateral areas and less so more medially. On this view, which we elaborate later, the amygdala would be just one of a variety of sites at which benzodiazepines act to reduce input to the behavioural inhibition system.

6.3.6 **Fear-potentiated startle**

The phenomenon of fear-potentiated startle requires special consideration, since it raises particular problems for the analysis of anxiety pursued in this book. Its importance rests on four factors: (a) the extensive psychological analysis to which it has been subjected; (b) the extensive detail in which the pathways involved have been elucidated; (c) its extensive pharmacological analysis; and (d) its demonstrated relation to state anxiety in human beings and to anxiety disorder (e.g. Grillon *et al.* 1993). The key facts are reviewed by Davis (1992b; Davis *et al.* 1993); where references are not given for statements in the rest of this section, they will be found in his reviews.

The paradigm itself is simple. A rat is first exposed to pairings of a light with a shock in a conventional classical conditioning arrangement. After these training trials, a loud noise is presented (in the dark) which results in a startle response. Finally, the loud noise is again presented but in the presence of the light and at about the time that the shock would have been delivered. A much larger (potentiated) startle response is produced as a result of the prior fear conditioning compared to the lesser response which might have been expected if habituation to the noise stimulus had occurred. A range of tests suggests that potentiation results from true fear conditioning and not from postural adjustments and other secondary effects (Davis 1992a, p. 257). An equivalent result can be obtained using the eyeblink to measure startle in human subjects.

The neural pathways involved have been worked out in considerable detail and conform tightly to the general plan described above. The light CS is received by the retina and relayed to the lateral and basolateral nuclei of the amygdala via the perirhinal cortex. Note, however, that 'complete removal of all primary and secondary visual cortices does not block the expression of fear potentiated startle using a visual conditioned stimulus' (Davis 1992a, p. 274), suggesting that subcortical 'quick and dirty' input is adequate to

support conditioning in the absence of a clear cortical signal. The unconditioned stimulus (shock) is detected by the footpads and relayed in the spinal cord and some unknown intermediate area before arriving in the basal nucleus. Here the conditioning of the CS by the UCS depends on an NMDA receptor-dependent process (see Davis *et al.* 1993, p. 190), presumably long-term potentiation. The information from the lateral and basal nuclei is then transferred to the central nucleus; this makes a monosynaptic connection with the nucleus reticularis pontis caudalis, which constitutes 'an obligatory part of the acoustic startle pathway' (Davis 1992b, p. 263; Hitchcock and Davis 1991; the full details of the startle pathway are given by Davis 1992a; see also Frankland and Yeomans 1995). Similar organization is found when an auditory stimulus is used for the conditioning phase, with perirhinal cortex again normally (but not necessarily) being involved (Campeau and Davis 1995b).

Campeau and Davis (1995a) trained individual rats with both auditory and visual conditioned stimuli and demonstrated that conditioning to one stimulus was independent of conditioning to the other. They then demonstrated that neurotoxic lesions of the central nucleus or of the basolateral complex essentially eliminated fear-potentiated startle equally for visual and for auditory conditioned stimuli (with a small residual amount of fear conditioning in the case of pre-training lesions of the basolateral complex). They further showed that the central nucleus lesions 'destroyed neurons, including those in the medial aspect of the central nucleus, which directly innervated the acoustic startle pathway' (Campeau and Davis 1995a, p. 2309; see Rosen *et al.* 1991 for the relevant anatomy). Taken together with the known anatomy of the amygdala and the data indicating similar results with other types of conditioned fear response, this suggests that 'in fear conditioning the basolateral complex of the amygdala serves as an obligatory relay of sensory information from subcortical and cortical sensory areas to the central nucleus of the amygdala . . . [and] that the central nucleus serves as a response-independent, final common relay for fear conditioning' (Campeau and Davis 1995a, p. 2301). This suggestion is consistent with the fact that the *sensitization* of fear-potentiated startle (its increase by co-presentation of other stimuli) also depends on the central nucleus, which in this case appears to be relaying input from locus coeruleus rather than the basolateral amygdala (Fendt *et al.* 1994). However, while the central nucleus is normally the basis for fear-potentiated startle, it is not absolutely necessary. Provided lesions are restricted to the central nucleus, animals can re-acquire the response post-lesion (Falls and Davis 1995).

While the paradigm is universally referred to as *fear*-potentiated startle, Davis links it with anxiety in a number of ways. First, he notes that potentiated startle indicates anticipation of the shock. Second, he links the increased startle response with equivalent changes in heart rate, salivation, ulceration, respiration, scanning and vigilance, urination, defecation, grooming, and freezing—all of which have corresponding features in generalized anxiety disorder. Third, and most critically (as we noted earlier), fear-potentiated startle is blocked by the majority of anxiolytic drugs (ethanol, barbiturates, benzodiazepines, buspirone, propranolol). Given the route we have adopted of seeking to define the neuropsychology of anxiety by reference to the action of these drugs, we too are required by the logic of our argument to accept fear-potentiated startle as a valid animal model of anxiety. Indeed, this phenomenon is virtually paradigmatic for the

'increased arousal' output of the behavioural inhibition system (see, for example, Gray 1987b). But—and we come now to the problem posed by fear-potentiated startle for our theory—we have seen that this behaviour is critically dependent upon the integrity of the amygdala. Furthermore, it is *not* dependent upon the integrity of the septo-hippocampal system. These considerations based upon fear-potentiated startle are backed up by other data on the increased arousal which is a feature of anxiety. As already indicated in the first edition of this book (Gray 1982, p. 182; and see Section 6.6 below), there is no evidence that the septo-hippocampal system mediates the increased arousal output of the behavioural inhibition system, and some that it does not. The data summarized in the present section clearly implicate the amygdala in this output. Here, then, is one part of the behavioural inhibition system whose neural basis must be located in the amygdala, not in the septo-hippocampal system.

While these data on fear-potentiated startle increase the neural complexity of our overall theory, they do not invalidate it. It was clearly stated in the 1982 edition that the theory does not identify the behavioural inhibition system with the septo-hippocampal system: the behavioural inhibition system requires the function of structures additional to the septo-hippocampal system, and the septo-hippocampal system has functions additional to those that it discharges in states of anxiety. With regard specifically to the arousal output of the behavioural inhibition system, having failed to locate this within the septo-hippocampal system, we tentatively attributed it in the first edition to the noradrenergic innervation of the hypothalamus (Gray 1982, p. 358). While this pathway may indeed play such a role, the extensive work on fear-potentiated startle summarized above indicates that this is likely to be minor compared to that of the amygdala. But this change in the identification of the neural basis, already accepted as being extra-hippocampal, of the arousal output of the behavioural inhibition system does not itself affect the fundamentals of the theory. Taken together, however, with other features of the theory, the location of the arousal output of the behavioural inhibition system, at least as this is manifest in fear-potentiated startle, does raise two issues that require consideration.

The first concerns the nature of the behaviour measured in the fear-potentiated startle paradigm. In this second edition of the book, we have placed greater emphasis on the role of conflict in defining the type of situation which gives rise to anxiety. In the first edition, in contrast, it was usually regarded as sufficient that a form of behaviour be elicited by a conditioned aversive stimulus for it to count as being related to anxiety (see, however, the discussion of on- vs. off-the-baseline conditioned suppression in Gray 1982, p. 296). The stronger emphasis on conflict, however, raises the question: what is the source, if any, of conflict in fear-potentiated startle? At first sight, there is none: all that happens is that a light (threatening shock) is followed by the startle-eliciting auditory stimulus—the animal has no (explicit) motor program with which this sequence of events can come into conflict. To resolve this problem, note that the increase in arousal demonstrated by the potentiated startle response is elicited by the light, not by the loud noise that actually triggers it. The latter stimulus is necessary for the paradigm only in providing a conveniently observable measure, startle, of the increased arousal. At the time of this increased arousal (during the interval between presentation of the light and noise), the animal is usually immobile; and, since the shock is inescapable, is likely to

be engaged in an irresolvable conflict between a variety of different potential escape responses, all of which have proved equally useless in the past. There is experimental evidence that, as implied by this analysis, fear-potentiated startle is the result of processes intimately related also to behavioural inhibition. Leaton and Borszcz (1985, p. 421) found that 'startle amplitude was positively correlated with freezing [during the CS] under all conditions [although there was] . . . a non-monotonic relation between potentiated startle and shock intensity'. Since conditioned freezing is sensitive to anxiolytic drugs, these data suggest, then, that potentiated startle is linked to anxiolytic-sensitive behavioural inhibition. On this view, the paradigm is one of anxiety-potentiated startle, not fear-potentiated startle. Crucially, the fact that anxiolytic drugs do not affect one-way active avoidance, and actually improve two-way active avoidance and Sidman avoidance (Chapter 4), shows that their reduction of fear-potentiated startle cannot be the result of a reduction in fear itself; and, conversely, that the potentiated startle phenomenon does not contribute to learned avoidance.

The second issue concerns the site of action of the anxiolytic drugs. Having defined first anxiety and then its neuropsychology in terms of the action of these compounds, we need to ask how it is that they come to pick out just those activated structures whose activity underlies anxiety. In the first edition of this book, we found a unifying principle able to offer an answer to this question in the postulate that the anxiolytics act principally by reducing stress-induced increases in the activity of the ascending monoaminergic pathways which innervate the septo-hippocampal system (a matter we consider in detail later). By locating the arousal output of the behavioural inhibition system in the hypothalamus, but secondary to an increased noradrenergic input to this structure, we were able to retain the simplicity of this unifying principle. Locating it, however, as we do now, in the amygdala, we lose this simplicity. This is a problem to which we return later in the book (Chapter 11, Section 11.3).

6.3.7 **Autonomic signs of anxiety**

Alongside this loss of simplicity on one front, however, the neuropsychology of the arousal output of the behavioural inhibition system to which the data on fear-potentiated startle conduct us brings a balancing gain of simplicity on another. At the time of writing the first edition of this book there was no clear picture of the neural structures that might mediate the autonomic symptoms of anxiety (including, for example, raised skin conductance, increased spontaneous fluctuations in skin conductance, raised heart rate, increased forearm blood flow; Lader 1975); although we suggested (Gray 1982, p. 362) that descending fibres from the locus coeruleus might be involved. As in the case of the increased arousal measured in fear-potentiated startle, Davis's (1992b) work clearly indicates that the amygdala (and especially the central nucleus) provides a stronger candidate; although descending fibres from the locus coeruleus may play an additional role. As summarized in Fig. 6.3, based on Davis's work, the central nucleus of the amygdala provides outputs to a whole range of lower structures which between them organize the entire spectrum of autonomic responses that occur during anxiety. Our balancing simplicity, therefore, is now to find in the amygdala a high-level control centre for both the arousal and the autonomic changes seen in anxiety.

Note, however, that these autonomic changes are not specific to anxiety: they occur during other types of defensive behaviour as well. Indeed, it is generally the case that (contrary to the well-known James–Lange theory of emotion) there is very little specificity in the relationship between changes in the autonomic nervous system and particular states of emotion (Gray 1987b, Chapter 5). Thus the same autonomic changes are associated with activity in both the fight–flight and behavioural inhibition system (and correspondingly in human states of both panic and anxiety, albeit with greater intensity in the former case; Lader and Mathews 1970). An appealing way of putting these sets of data together, therefore, is to consider the interactions between the behavioural inhibition and fight–flight systems as being in part mutually inhibitory and in part mutually reinforcing. In particular, we suggest that that part of anxiety which consists in heightened arousal and increased activity in the autonomic nervous system is shared by the fight–flight and behavioural inhibition systems; but that the behavioural inhibition output of the behavioural inhibition system is exercised (*inter alia*) on the fight and flight outputs of the fight–flight system. This arrangement makes good adaptive sense. In a situation of potential threat (giving rise *ex hypothesi* to anxiety) the requirement for an increment in arousal lies precisely in the possibility that the animal will need to engage in the vigorous behaviour that constitutes fight or flight. Similarly, the changes that take place in the autonomic nervous system have the function of increasing the body's readiness for, and ability to sustain, such vigorous action (Gray 1987b). The easiest way for the behavioural inhibition system to produce these arousal and autonomic outputs, therefore, is simply to have the fight–flight system simultaneously activated. At the same time, however, the behavioural inhibition system needs to restrain the *behavioural* outputs of the fight–flight system (fight and/or flight) until they are most adaptive. (When we come, in Chapter 11, to consider the application of our model to clinical phenomena, this postulate will help us to understand the otherwise curious fact that certain anxiety-related disorders, notably panic and post-traumatic stress disorder, are associated with evidence of damage to the hippocampal formation.)

There is an additional likely reason for separate control of the arousal/autonomic and behavioural inhibition outputs of the behavioural inhibition system. Our analysis of the interactions between the fight–flight and behavioural inhibition systems suggests that these can often be co-activated, with a need for very sudden switching between them because of their essentially complementary but opposing roles in directing the animal's behaviour. Autonomic changes, by contrast, are relatively slow and required to a similar extent whether it is the fight–flight or the behavioural inhibition system that is at any given time in control of behaviour.

6.4 THE DEFENCE SYSTEM—INTEGRATION

We have now covered all but the highest levels of Table 6.1. This table is, of course, too categorical, too 'cut and dried', given the extensive reciprocal connections between adjacent levels. However, we believe it to be correct in principle and so leave it as it is for the sake of simplicity.

As detailed in Table 6.1, defensive systems appear to be organized hierarchically. The periaqueductal grey is at a low level and is concerned with the release of a suite of motor

programs (undirected escape, aggression) designed to remove the animal from harm as quickly as possible and with the minimum of analytical finesse. To adapt LeDoux's comment on low-level visual input to the amygdala, the periaqueductal grey can provide an ultra-quick, ultra-dirty solution to problems of defence. The medial hypothalamus is at the next level and can be viewed as performing essentially the same function (e.g. directed or learned escape, while now inhibiting aggression), but under conditions in which there is more time, more flexibility in terms of options, and where conditioning can influence responding. Despite being (or indeed because of being) at the level above the periaqueductal grey, the medial hypothalamus can be viewed as operating via connections with the latter region (some excitatory, some inhibitory). The amygdala is at the next level again. Here complex conditioning is possible (e.g. active anticipatory avoidance). In addition, the amygdala mediates responses to (unconditioned) omission of reward (so, for example, amygdalar lesions disrupt the frustration effect in the double runway; Henke 1977). Like the hypothalamus, to discharge these functions the amygdala must have both excitatory and inhibitory connections with lower levels. The final level, at which active avoidance involves more complex stimuli, both learned and innate, is the anterior cingulate cortex, which is considered below.

The picture we have presented of threat analysis by the amygdala (following closely LeDoux 1994) is of a structure which receives, from both higher levels of sensory cortex and lower levels of subcortical sensory afferent systems, information about the presence of threatening stimuli, about novel stimuli, and about previously neutral stimuli occurring in the context of threat. Aversive conditioning then occurs in the amygdala through long-term potentiation of synapses linking incoming, previously ineffective, stimuli with the defence control system. The output from the amygdala then results in output from the hypothalamus and periaqueductal grey, if the simplest types of autonomic and motor response are sufficient; or from the more general motor programming areas of the cortex and basal ganglia if more flexible avoidance is required. Note that the associative nature of long-term potentiation is consistent with an architecture in which the amygdala controls, via input from relevant innate stimuli, reactions that can become conditioned to neutral stimuli.

In what we have described so far, we have touched only in passing on the matter of how the various elements of this defensive system operate together. Since they all react to threat, this might not seem a problem. However, undirected escape, directed escape, and avoidance can involve mutually incompatible responses and strategies. How does the brain deal with this problem?

To provide a preliminary answer to this question, we return to Graeff's (1994) review with which we started this chapter. The basic principles, if not the complete picture, of the mutual organization of the different levels of the defence system are exemplified by the effects of serotonergic (5HT) input from the raphe onto this system.

> Both facilitatory and inhibitory functions have been ascribed to 5HT. The . . . view that 5HT enhances anxiety is mainly based on the observed effects of drugs and neurotoxic lesions in animals submitted to conflict tests. In general, operations that decrease 5HT activity in the brain release behaviour suppressed by punishment in the rat. A likely site for this facilitatory action of 5HT is the amygdala, since it has been shown that microinjection of 5HT receptor antagonists inside the

basolateral amygdala has anxiolytic effects, whereas 5HT receptor agonists enhance behavioural suppression determined by punishment.

In contrast, the results of studies using escape from electrical stimulation of the dorsal midbrain of the rat led to the view that 5HT inhibits defensive behavior in the dorsal PAG [periaqueductal grey]. For example, it has been shown that systemically administered drugs that enhance 5HT activity have anti-aversive effects while those that impair 5HT transmission have pro-aversive effects. . . .

Taken together, the evidence reviewed above suggests that 5HT facilitates defensive behavior integrated in the amygdala, while inhibiting the defence reaction organized in the PAG. In order to make sense of this seemingly paradoxical situation, Deakin and Graeff (1991) hypothesized that the PAG would be responsible for proximal defence, commanding fixed action patterns that are only adaptive under conditions of extreme danger. Conversely, if the same behaviors were performed in response to potential or distal danger they would enhance the probability of detection by an aggressor. Instead, more flexible, largely learned behavioral strategies are likely to lead to successful escape from or avoidance of danger. . . . Translating it into emotional terms, it may be said that 5HT enhances anxiety and fear while at the same time inhibiting panic. (Graeff 1994, pp. 820, 821.)

Deakin and Graeff's (1991) model is shown, with some modifications, in Fig. 6.4. It can be seen that the addition of the 5HT system, essentially in parallel with the hierarchical defence control system, provides at least some of the required coordination of

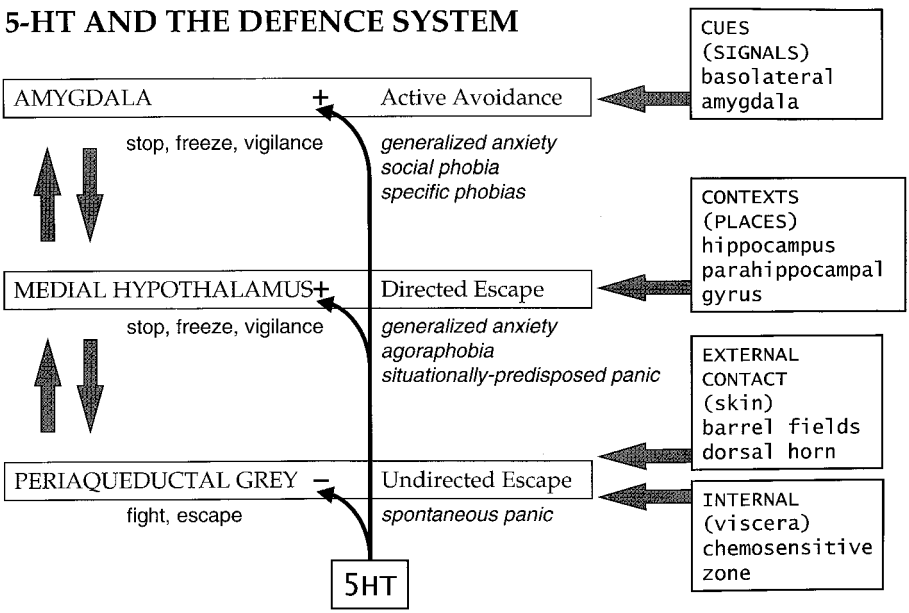


Fig. 6.4 The role of 5-hydroxytryptamine (5HT) in the control of the defence system proposed by Deakin and Graeff (1991). The key feature of their proposal for our purposes is the opposing regulation of the periaqueductal grey and more central structures and the specific role of 5HT in this control. This structuring allows one to account for phenomena such as relaxation-induced panic. For further discussion, see text.

the different levels. This model may also 'explain why the $5HT_2$ receptor antagonist ritanserin has been shown to ameliorate anxiety in patients with a diagnosis of generalized anxiety disorder, whereas panic disorder patients were found to be either not improved or even worsened by the same drug' (Graeff 1994, p. 821). Thus, Graeff's view of the raphe 5HT system is of one which carries a simple signal that an *avoidable* danger is present. This both activates those areas concerned with avoidance and inhibits those areas which would be likely to produce interfering escape responses.

Similarly, as already noted, activation of the medial hypothalamus and periaqueductal grey, respectively, elicits very similar escape and avoidance responses (better coordinated, however, at the hypothalamic level). However, these two structures have opposing influences on aggression, an arrangement which increases the probability that aggression will occur in the context of uncoordinated escape (with which it is functionally synergistic) while decreasing the possibility that aggression will occur in the context of coordinated escape (with which it will interfere). There are likely to be many more such mutually modulatory interactions between specific components of specific nuclei of the subcortical defence system.

6.5 INTERIM CONCLUSIONS

We have presented a strongly hierarchical view of the defence control system, modelled largely on the ideas presented in Graeff (1994; but see also Sudré *et al.* 1993). The periaqueductal grey, medial hypothalamus, and amygdala control uncoordinated escape, coordinated escape, and avoidance respectively. The requirements of these different responses vary and their interaction is coordinated by interactions between the different levels of the defence system and by the ascending serotonergic system. An important point, to which we return in Chapter 11, is that the hierarchical organization of the system in terms of sophistication of analysis and control nonetheless means that the lowest level of our analysis, the periaqueductal grey, can be viewed as the main output station for primary defensive reactions arising at all levels (e.g. Blanchard *et al.* 1981; see also the specific model presented by Fanselow 1991).

A particularly pleasing picture of the contribution of the amygdala to defence is provided by combining the ideas of Davis (1992a) with those of LeDoux (1994, 1996). The amygdala receives relatively low-grade, but affectively labelled, information about sensory inputs from the thalamus. It receives (with somewhat greater delay) extremely highly processed, cognitive rather than affective, information about the same sensory inputs from the highest levels of unimodal and polymodal cortex. Novel sensory stimuli are sent to the lateral nucleus of the amygdala but, unless they are associated with affective input, they rapidly habituate. Learned avoidance depends on the pairing of a CS (arriving in the lateral nucleus) with a UCS (arriving in the basal nucleus) and on the resultant long-term potentiation of the connection between the CS and the UCR. This is the paradigm case of Pavlovian stimulus substitution. In principle, specific biologically prepared stimuli (Chapter 11, Section 11.13) could operate in a similar fashion with their initial input being sufficient, even in the absence of long-term potentiation, to produce a UCR. The various outputs of the central nucleus of the amygdala then control

the different components of different UCRs and CRs. Fear would, on this view, result from activity in the basal nucleus of the amygdala and this could lead to: (a) defensive reactions via the subcortical outflow from the central nucleus; (b) the adjustment of learned avoidance via the outflow from the basal nucleus to the striatum; (c) anxiety via the outflow to the septo-hippocampal system; and (d) focusing of attention and increased negative cognitive bias via the recursive outflow from the basal nucleus to all levels of the visual cortex and other sensory systems both directly and via the basal forebrain cholinergic system (see Appendices 2 and 10).

So far, then, we have dealt with brain defence systems which provide precursors to anxiety but which are only partially involved in anxiety itself. We next move on to consideration of the septo-hippocampal system. Like Graeff (1994), we view this system as one of the higher levels of the defence control system; and, like LeDoux (1994), we view it as one of the higher levels of the threat analysis system. On either view, the septo-hippocampal system is likely to have a major role to play in the elaboration of anxiety. However, in distinction to both Graeff and LeDoux we view this system as having an antithetical, or at least doubly dissociable (Selden *et al.* 1991), role in comparison to the amygdala, resulting from the fact that fear and anxiety have, in many cases, opposite outputs and, in many other cases, do not overlap (see also Bechara *et al.* 1995).

6.6 COMPARISON BETWEEN THE SEPTAL, HIPPOCAMPAL, AND ANXIOLYTIC SYNDROMES

We leave a detailed consideration of the operation and functions of the septo-hippocampal system to Chapters 7–10 and to Appendices 4–9. For the purposes of the present argument, we note only that there is extensive evidence that anxiolytic drugs produce rather subtle impairments in the functioning of the septo-hippocampal system; and that such impairments, when produced selectively, change behaviour in ways common to the effects of both systemic administration of the anxiolytic drugs and septo-hippocampal lesions (see Section 1.5 in Chapter 1). The next step, addressed here with lesion data, is to choose between two possibilities: (1) that the hippocampus represents a source of highly processed cortical information, which then allows the amygdala to generate anxiety in certain cognitively complex situations (e.g. as often suggested, when anxiety must be conditioned to contextual as opposed to simple stimuli); or (2) that the hippocampus itself generates anxiety, interacting with the amygdala when anxiety depends on fear stimuli and when the level of anxiety is such that autonomic activation is required (see Section 6.3.7, above).

Table 4.2 summarizes the septal and hippocampal syndromes described in detail in Appendix 8. It should be noted that (a) this table is based on large lesions of the septum and relatively large lesions of the hippocampus, both of which are likely to have interfered with fibres of passage; (b) our major concern here is to define a 'septo-hippocampal syndrome' which is based on the overall points of agreement between the two types of lesion; and (c) we have excluded the extensive literature analysing the involvement of the hippocampus in memory, because there are few

equivalent data on anxiolytic drugs. This omission is redressed in Chapter 8, where we consider the analysis of specific types of memory task in relation to hippocampal function.

Our reasons for comparing septal with hippocampal lesions are that there are extensive interconnections between the two areas (described in detail in Appendix 4; see Fig. 1.3), and that the septum controls hippocampal theta, through which we postulate that anxiolytic drugs produce critical behavioural effects. It is, therefore, not surprising that the septal and hippocampal syndromes should resemble each other to some extent. However, in view of the pitfalls of the lesion technique and the extensive connections of the hippocampus to the cortex, it is perhaps surprising that the effects of septal and hippocampal lesions resemble each other so closely (Table 4.2). The cases where septal and hippocampal lesions have different effects from each other are discussed in detail in Gray (1982, Section 6.24). Here we are concerned only with the cases where the effects of the two kinds of lesion are concordant.

Of course, the 'septo-hippocampal syndrome' so defined may in part be fortuitous, or due to interruption by septal lesions of hippocampal afferents or efferents which do not relay in the septum (as has been shown for the effects of septal lesions on latent inhibition and the partial reinforcement extinction effect; see Appendix 9). However, for our present purposes, this syndrome is only a stepping-stone to more detailed analysis, and a high degree of anatomical precision is not essential. The same is true of our behavioural categories. Throughout our review we have found that the presence or absence of a deficit in any nominal behavioural task can depend on the presence or absence of quite detailed features of the procedure or of the animal's inherent response tendencies. Nonetheless, in the absence of similarly detailed analysis of all the treatments to be compared, we must take a coarse-grained average and compare groups of tasks in terms of whether an effect or lack of effect is most likely to be observed for that particular type of task.

The anatomical interconnections between the septal area and hippocampus provide strong a priori grounds for expecting the septal and hippocampal syndromes to resemble each other, irrespective of any particular theory of their behavioural functions. However, there are no a priori grounds, other than those derived from our theory, for expecting any similarity between the septo-hippocampal syndrome and the effects of anxiolytic drugs. But it is clear when we compare the third and fourth columns of Table 4.2 that the concordance here is also remarkably strong. Since the value of science is often held to reside in its predictive power, it should be noted that the overwhelming majority of the data which support the concordance in Table 4.2 were collected *after* the original proposal of the links between anxiolytic drug effects and septal and hippocampal lesions (Gray 1970a).

The degree of concordance between the anxiolytic and septo-hippocampal syndromes is particularly surprising in the case of benzodiazepines, because their receptors are widely distributed throughout the brain. Their muscle-relaxant influence, at least, might have been expected to produce many effects which would decrease the concordance with the septo-hippocampal syndrome. However, where it has been tested in the tasks we have considered, buspirone almost invariably shares the effects of the benzodiazepines and hence of septal and hippocampal lesions. This suggests that the battery of tests we have

reviewed are not, by and large, sensitive to the muscle relaxant, anticonvulsant, hypnotic, depressant, and addictive properties of the benzodiazepines—since buspirone's side-effects do not include any of these properties. The striking feature of Table 4.2 is that, with few exceptions, wherever the effects of septal and hippocampal lesions are the same as each other, the effects of the anxiolytic drugs are the same (although there are many cases where relevant data are not available). The size of the drug effect, however, is usually smaller than that of the lesions.

An exception to the rule of anxiolytic/septo-hippocampal similarity is found in aggressive behaviour. Although the effects of the drugs on aggression are very variable (Gray 1977), there is reason to suppose that there is a real divergence on this measure between their effects and those of septal and hippocampal lesions. Both septal and hippocampal lesions usually reduce aggression if this is induced by means other than shock (usually by isolating animals). When shock is not used, all the anxiolytic drugs have been reported to increase aggression, particularly at low doses. As the blank spaces of Table 4.2 become filled, we may expect more cases like this.

There are also few discrepancies in the reverse direction. Those that exist can be attributed to one effect of anxiolytic drugs which, as noted above in Section 6.3.6, appears to be missing from the septo-hippocampal syndrome, and which is of particular importance for the psychological theory developed in Chapter 5. The drugs impair all three postulated outputs of the behavioural inhibition system: behavioural inhibition itself, increased attention to novel stimuli, and increased level of arousal. However, while the septo-hippocampal syndrome includes the first two of these, there is no evidence that septo-hippocampal lesions lower the level of arousal.

Several experiments suggest more definitely that they do not have this effect. Dickinson (1975) trained rats on a discrete-trial task in which lever presses were rewarded with food on a variable ratio (VR) 2 schedule (equivalent to a random 50 per cent partial reinforcement schedule). Shock was delivered for lever pressing on the same schedule. In one condition it was perfectly positively correlated with food delivery, in another negatively. In control animals, shock suppressed lever pressing more in the negatively than in the positively correlated condition. This can be attributed to counter-conditioning of the effects of shock by the immediately following food in the positively correlated condition (Dickinson and Pearce 1977). In fact, in the positively correlated condition, lever pressing was not suppressed in control animals at all. In both the positively and negatively correlated conditions, response rate was increased by septal lesions. In the negatively correlated condition, there was less suppression than among controls, as would be expected. In the positively correlated condition there was an actual *increase* in response rate related to the *unsuppressed* control baseline. One interpretation of these findings (Dickinson 1974, 1975; Gray and Smith 1969) is that the level of suppression seen in the controls represents an algebraic summation of two separate effects of punishment: a rate-decreasing effect (behavioural inhibition proper) and a rate-increasing effect (increased arousal, or what in Hullian theory was known as drive summation; see Gray 1975), and that septal lesions eliminated the former but not the latter.

A second experiment which lends itself to this explanation is a study of the effects of distraction on the behaviour of rats in the alley. Raphelson *et al.* (1965) showed that a

novel stimulus caused running speeds to decrease in normal animals, but to increase in animals with hippocampal lesions. Again, this result may reflect a loss after hippocampal lesions of behavioural inhibition, but with preservation of the increment in arousal caused by novel stimuli.

But perhaps the most significant result in this category is the fact that hippocampal lesions do not reduce fear-potentiated startle as do the anxiolytic drugs. As indicated above (Section 6.3.6), this effect of the anxiolytics is almost certainly achieved through a direct action on the amygdala. This pattern of findings suggests that the same may be true for all of the cases in which anxiolytics affect arousal and septo-hippocampal damage does not.

Overall, then, inspection of Table 4.2 shows that the similarities between the effects of septo-hippocampal damage and anxiolytic drugs, in terms of decreases, increases, and the equally important cases of no change, cover a very wide range of behaviour. The lesions share the pattern of the drugs' effects in tasks that seem to rule out any general effects on memory as opposed to emotion (for example, passive avoidance, impaired, compared to one-way active avoidance, unchanged). Similarly, the drugs share the lesions' effects on tasks which seem to involve memory more than emotion (water maze, delayed conditional discrimination). In addition there is one similarity which is less well established: under certain conditions all three treatments reduce the animal's capacity to attend to and take in information about novel features of its environment. Taken together, these similarities offer strong support for the hypothesis that the anxiolytic drugs produce at least some of their effects by impairing the functioning of the septo-hippocampal system, directly or indirectly.

At the start of this section we distinguished between two possible relationships between the septo-hippocampal system and the amygdala. The data offer little support for the first possibility, that the hippocampus provides to the amygdala a specific subset (usually taken to be spatial or contextual) of the available stimuli from which the amygdala then constructs anxiety. If this hypothesis were correct, amygdalar lesions should produce effects that always (or nearly always) overlap the anxiolytic profile, whereas hippocampal lesions should do this only in tasks involving the relevant stimulus subset. The actual pattern of discrepancies between the three profiles does not fit this pattern. As can be seen from Table 4.2, and reiterating the summary already presented in Section 6.3.2, the effects of amygdalar lesions are both positively and negatively discrepant from the anxiolytic/septo-hippocampal syndrome. Lesions of the amygdala decrease one-way active avoidance, decrease two-way active avoidance, increase rearing, and decrease the frustration effect. In all of these cases anxiolytic drugs and septal and hippocampal lesions are without effect or have the opposite effect. Conversely, lesions of the amygdala are without effect on behaviour in most delayed matching to sample tasks and on fixed interval schedules. Both of these are affected by anxiolytic drugs and by septo-hippocampal lesions. The latter pattern is consistent with the second of the two possibilities distinguished at the start of this section, namely, that the hippocampus generates anxiety, interacting with the amygdala only when anxiety either depends on fear stimuli (not involved, for example, in fixed interval performance) or when the level of anxiety is sufficiently high to require autonomic activation (unlikely, for example, for delayed matching to sample with normal levels of experimental food deprivation).

Overall, then, we can conclude that the amygdala is a poorer candidate for the 'seat of anxiety' than is the septo-hippocampal system. Such similarity as lesions of the amygdala show to anxiolytic drugs in their effects in conflict situations are best attributed to reductions in fear, which weaken the avoidance component of the approach–avoidance conflict. Thus, amygdalar lesions impair passive avoidance (and can be held to do so by altering fear in the same way that accounts for their effects on active avoidance), but do not impair conflict in tasks such as fixed interval rewarded bar-pressing, where frustration rather than fear is the source of the conflict. However, this is not to say that the amygdala has no role in anxiolytic drug action. As we saw in Section 6.3.6, the effects of the drugs on fear-potentiated startle (and perhaps on all comparable 'increased arousal' effects) are mediated via the amygdala.

6.7 THE EFFECTS OF LESIONS OF THE NORADRENERGIC AND SEROTONERGIC SYSTEMS

We have already briefly discussed Graeff's suggestion that the serotonergic system is one of the means through which activity in different levels of the defence system is coordinated. Noradrenergic systems have also in the past been suggested to be crucial for defence in general and anxiety in particular (see McNaughton and Mason 1980). These monoamine systems also innervate all of the components of the defence hierarchy considered so far (as well as innervating the frontal and cingulate cortices which we consider shortly; see Figs 6.5 and 6.6). It would not be surprising, then, if the monoamines were crucial for at least some aspects of anxiety. As in the previous sections, we shall assess their possible role via the effects of lesions.

Table 4.2 is regrettably empty in the case of serotonergic systems but satisfactorily full for noradrenergic systems. The pathway most studied (and to which we confine our review) consists of the dorsal ascending noradrenergic bundle, which carries afferents to the whole of the forebrain, including the septo-hippocampal system, from the locus coeruleus. Briefly, where lesions of the dorsal bundle have behavioural effects, these usually concur with those making up the anxiolytic/septo-hippocampal syndrome; but there are a large number of cases where dorsal bundle lesions do not reproduce the effects of anxiolytic drugs (Appendix 10). Interestingly, so far as can be told from the small amount of data, the effects of serotonergic lesions are to a large extent complementary to those of damage to the dorsal noradrenergic bundle, that is, they appear to supply the missing elements of the anxiolytic/septo-hippocampal syndrome. (Note, however, that the effects of dorsal bundle lesions sometimes become evident only in adrenalectomized animals. There may, then, be considerable redundancy between these different stress-related systems.) The fractionation of the anxiolytic syndrome by lesions to the separate monoamine systems suggests that neither can be thought of as uniquely constituting an anxiety 'centre', although some common target (such as the amygdala and/or the septo-hippocampal system) might play that role.

It is consistent with this pattern of results that anxiolytic drugs act directly on the locus coeruleus and raphe nuclei (Appendix 10). Both noradrenergic and serotonergic fibres project to the septo-hippocampal system. A direct action of the anxiolytic drugs on the source nuclei should, therefore, have extensive effects on this system and, as a result, reproduce aspects of the septo-hippocampal syndrome. There is also evidence of synergistic interactions between the monoamine systems, implying that anxiolytics ought to have greater effects than would be predicted from their effects on the noradrenergic and serotonergic systems separately (Appendix 10). This said, we should note that the anxiolytic drugs have important actions on other inputs to the septo-hippocampal system (we discussed the cases of the amygdala above and of the supramammillary nucleus in Chapter 1, Section 1.5). Further, the serotonergic and noradrenergic systems project to many areas other than the septo-hippocampal system (e.g. the amygdala, frontal cortex, and cingulate cortex). On this basis the goodness of fit between the combined noradrenergic plus serotonergic behavioural profiles and the septo-hippocampal profile is surprising. We provide an account of this later, but to some extent it suggests that the septo-hippocampal system is unusually dependent on inputs from its monoaminergic afferents.

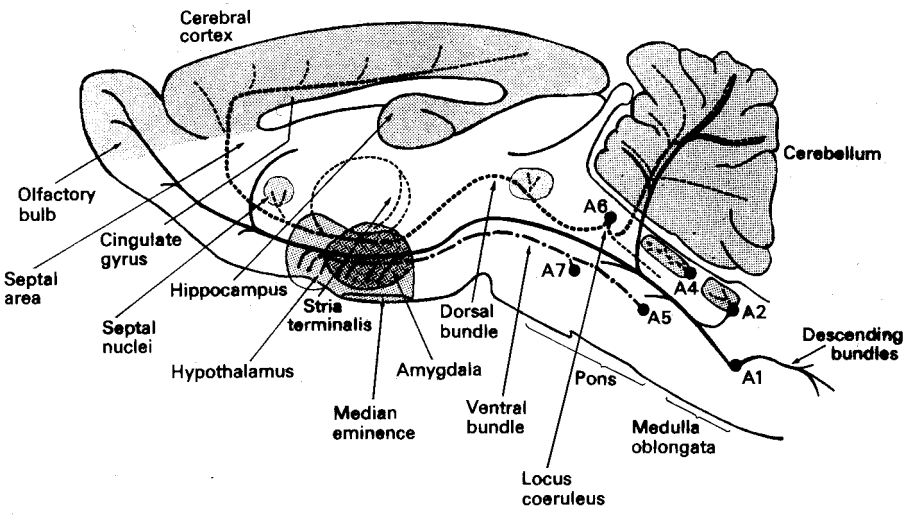


Fig. 6.5 Saggital representation of the rat brain, showing the principal ascending and descending noradrenergic pathways. Cell bodies in the locus coeruleus (A6) give rise to pathways (- - -) innervating all cortical areas of the brain. The dorsal bundle arising from A6 also innervates areas of the amygdala and hypothalamus. Shaded areas indicate regions of noradrenergic terminals. (From Livett 1973.) Note that ascending serotonergic innervation of the hippocampus follows a generally similar pattern to that of the dorsal bundle. A1, A2–A5, A7, other noradrenergic nuclei.

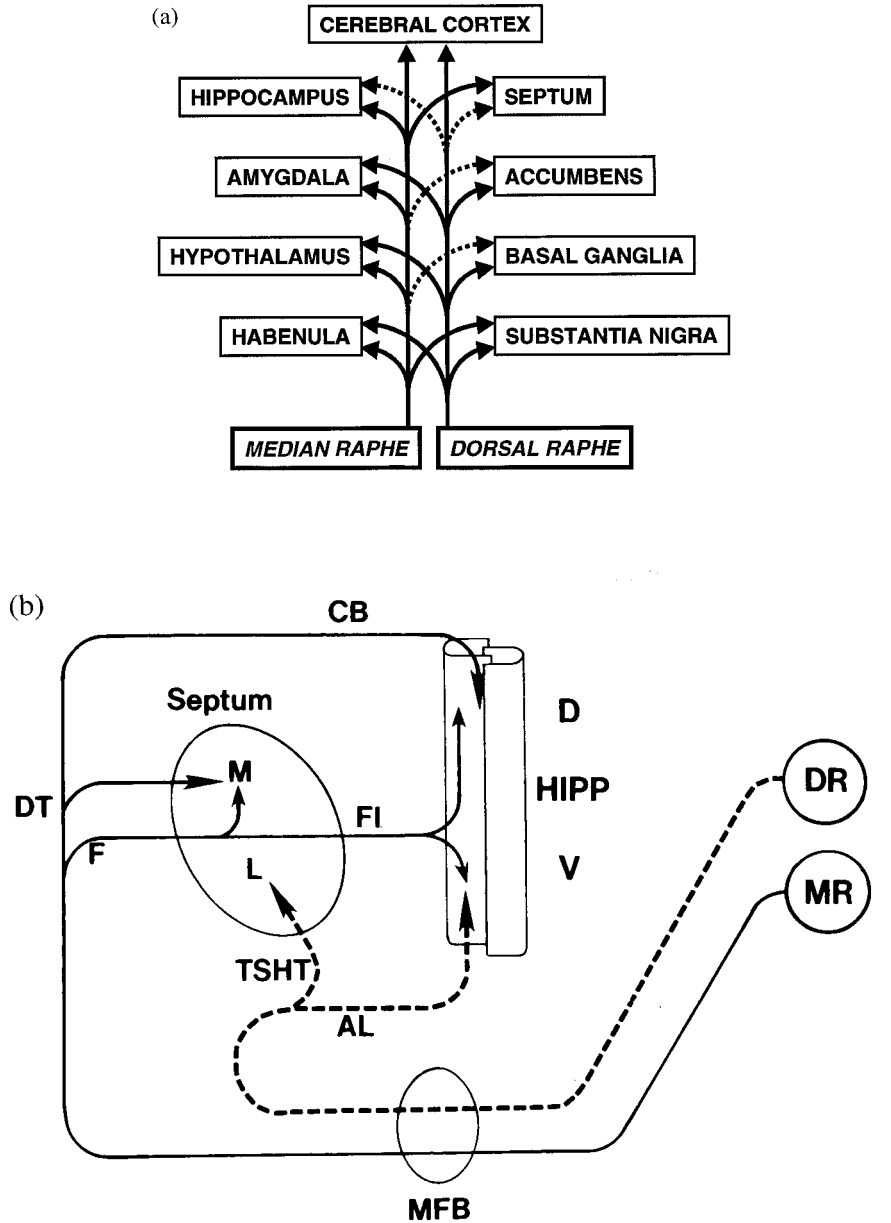


Fig. 6.6 Ascending serotonergic projections. (a) Schematic diagram of the major projection areas of the dorsal and median raphe. Dashed lines represent relatively weak connections. (Redrawn from Soubrié 1986.) (b) Diagrammatic representation of the main projections to the septo-hippocampal complex. AL, ansa lenticularis; CB, cingulum bundle; D, dorsal and V, ventral hippocampus; DR, dorsal raphe; DT, diagonal tract; F, fornix columns; FI, fimbria; L, lateral and M, medial septal area; MFB, medial forebrain bundle; MR, median raphe; TSHT, septo-hypothalamic tract. (From Azmitia, in Elliott and Whelan 1978.)

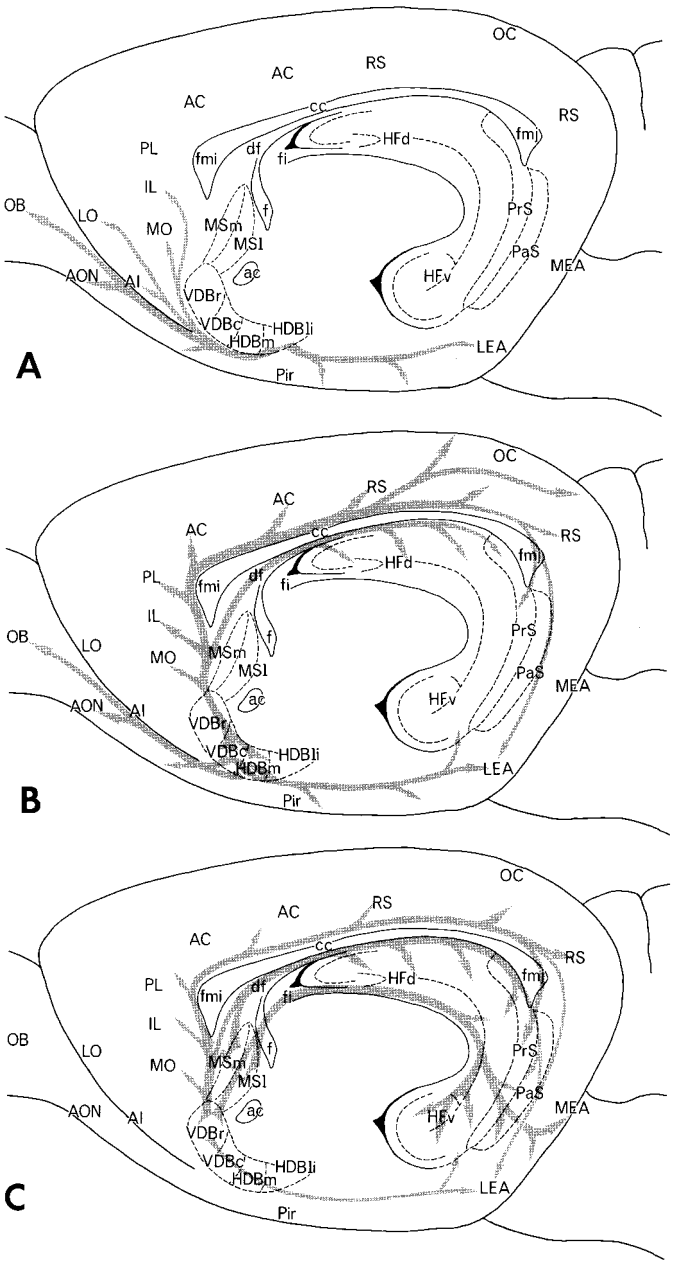


Fig. 6.7 Ascending cholinergic projections from the basal forebrain. These projections are topographically organized with: (A) relatively caudal basal forebrain areas projecting to more ventral areas of the forebrain; (B) relatively rostral areas projecting to more dorsal areas of the forebrain; and (C) relatively dorsal areas projecting relatively weakly to more dorsal areas of the forebrain and relatively strongly to the hippocampal formation. Note the tendency of fibres to follow the same routes as, and to have a similar innervation pattern overall to, the noradrenergic and serotonergic inputs to the forebrain. (From Gaykema 1992.)

6.8 THE CHOLINERGIC SYSTEM

Although somewhat different in general organization (and linked more often to memory than to emotion), the ascending cholinergic system has a distribution of efferents to cortex, hippocampus, and amygdala equivalent to those of the monoamine systems (Fig. 6.7) and, like them, is involved in the control of hippocampal theta activity. We need to consider, therefore, whether it too may make a contribution to the control of anxiety.

The data in the cholinergic column of Table 4.2 are not entirely satisfactory in that they have largely been provided by systemic injections of anticholinergics rather than specific neurotoxic lesions. However, if they can be taken at face value, they suggest a superficially surprising similarity to the anxiolytic drugs. The similarity with septo-hippocampal lesions is less surprising, given the extensive cholinergic innervation of the hippocampus and, as we discuss in Appendix 5, the crucial contribution of cholinergic input for the production of one type of hippocampal theta rhythm. Probably because anticholinergics are not used clinically as anxiolytics, drugs of this class have seldom been tested in the conventional anxiolytic screening tests. Where they have, they appear (unlike the data obtained in other types of experiment) *anxiogenic* rather than anxiolytic (Smythe *et al.* 1996). In contrast, largely because of the links between Alzheimer's disease, hippocampal dysfunction, and cholinergic neurodegeneration, anticholinergics have been extensively tested in memory paradigms. Given that anxiolytics have effects similar to hippocampal lesions on many memory tests, the resulting parallels in the table cease to be surprising.

We give this matter more attention later, but for the present we conclude that such similarity as there is between anticholinergics and anxiolytics is probably in large measure due to the fact that both classes of drug can act through the septo-hippocampal system. However, this overlap should be seen as at least partly accidental, since there are data suggesting that the qualitative nature of the drug effects is different even in tasks where they produce similar quantitative decrements (McNaughton and Morris 1987), and some types of memory task may be much more sensitive to anticholinergics than to anxiolytics (Kirk *et al.* 1988; Tan *et al.* 1996). Given the quite different patterns of impairment of theta control produced by the different classes of drug (Appendix 5), an incomplete overlap in their 'septo-hippocampal' behavioural effects is not surprising. Some non-septo-hippocampal parallels can also be expected, at least with benzodiazepines, as these compounds block release of cortical acetylcholine through an action in the basal forebrain, from which cholinergic neurons project to the neocortex (Moore *et al.* 1995; see also Holley *et al.* 1995; McGaughy *et al.* 1996).

6.9 THE PREFRONTAL CORTEX AND CINGULATE CORTEX

So far, as we have ascended the 'defence hierarchy', we have found a progressively better fit between the effects of lesions of individual structures and the effects of systemic injections of anxiolytic drugs. In the case of the septo-hippocampal system, the degree of fit (though still only partial) is truly remarkable, given the crudeness of the lesions involved. Simplifying somewhat, we can account for this partial overlap if we identify fear (elicited

by exposure to aversive stimuli without conflict) with activity in the amygdala, and anxiety (fear to which an approach-based conflict is added) with concurrent activity in the amygdala and septo-hippocampal system. The close anatomical relationship between these structures then makes good sense of the ethological and learning-theory relations we have already outlined between fear and anxiety.

If you view the neocortex as dealing with cognitive functions while subcortical areas deal with emotion and motivation, it might seem that the septo-hippocampal system (being a mixture of subcortical and archicortical structures) should be the apex of the defence system. However, a more evolutionary perspective suggests an alternative. With the unimodal sensory systems, phylogeny has added progressively more laminated neocortical areas to older systems to produce more sophisticated mechanisms for solving the same basic problems. These additions, however, do not fit our normal assumption that phylogenetically recent areas involve the most sophisticated processing. The most recent and anatomically sophisticated neural architecture is closest to the sensory receptors. Rather than being the site of the highest level of mental processing, it adds an extra, preliminary, filter to the existing system which then increases the level of processing in what are, in effect, the oldest areas. This can be seen as a developmental necessity, since each addition has occurred through the expansion and differentiation of the currently highest levels of the hierarchy. This must leave lower and/or earlier mechanisms, at least initially, largely unaffected so that they can continue to deal with those problems which they were originally capable of solving. (However, in the case of the sensory systems it is clear that this process has ultimately led to the encephalization of some functions so that, for example, the human tectum cannot perform, in the absence of the cortex, the kinds of task that are undertaken by the frog tectum.)

This perspective suggests that, even if the septo-hippocampal system is a key structure in the brain on which the processing of anxiety depends (as the superior colliculus and lateral geniculate are key structures on which vision depends), there could well be cortical areas which carry out more sophisticated forms of the same control of anxiety with which, on the lesion evidence, we would associate the septo-hippocampal system. There is a trivial sense, of course, in which all neocortical sensory systems (and ascending non-cortical sensory systems) must be involved in anxiety: they provide the sensory input and cognitive evaluation on which assessment of threat is based. This is the role that we (following LeDoux) have already assigned to them in relation to the conditioning of fear: sensory input is sent monosynaptically from different levels of the sensory systems to the amygdala, where it functions to provide neutral CSs until long-term potentiation (or some other form of plasticity) produces a functionally active connection of the input to defensive systems. These sensory areas must also provide the basis for the detection of, as opposed to reaction to, innate fear and innate anxiety stimuli. Similarly, all neocortical efferent systems (and descending non-cortical motor systems) could be involved, on occasion, in behavioural output resulting from anxiety. However, like the primarily sensory systems, this involvement is not specific to anxiety and can be linked to the motoric requirements of a specific situation rather than to its defensive aspects.

With the bulk of the cortex, then, lesion evidence provides us with little reason to suppose a primary involvement in anxiety. For example, after lesions of visual cortex, one can demonstrate that anxiety is intact, provided we use a different, intact sensory

modality to present significant stimuli, whereas all responses to visual stimuli, anxiogenic or not, are equally degraded. Similarly, one can show with lesions of, for example, motor cortex that there is a general loss of both non-anxious and anxious responses of some specific type, independent of eliciting stimuli. However, there are two cortical areas where lesion evidence suggests a more specific role in anxiety: the prefrontal and cingulate cortices. Damage to either of these regions has served successfully as a treatment of last resort for chronic, severe, and disabling states of anxiety (Marks, 1969; Powell 1979), particularly those characterized by obsessive–compulsive symptoms. We would need in any case to consider the prefrontal and cingulate cortices even when focusing on the hippocampus and amygdala, since they have close anatomical connections with both, and their outputs match those of the hippocampus much more than do those of any other neocortical areas.

Taken together, the prefrontal and cingulate cortex involve a huge part of the cortical mantle; have subdivisions which have been recognized only recently; and have been assessed functionally only with the coarsest of lesions and the most preliminary of field-specific single-cell analysis. In Appendix 3, therefore, we have attempted to build a picture based largely on neuroanatomy and speculation which attempts to make sense of, rather than being derived from, the lesion and single cell data.

The increasing cytoarchitectonic differentiation that occurs during evolution and development provides an important basis for simplifying and understanding the anatomical connections and functions of the frontal cortex. Cutting across this basic cytoarchitectonic and connectional organization is a division into cortex which has a 'hippocampal' and 'olfactory' origin, respectively. Thus one can see an evolutionary, cytoarchitectural, connectional, and functional 'dorsal trend' (see Barbas and Pandya 1987, 1991; also referred to as 'medial' by Goldberg 1985), originating in hippocampal allocortex and progressing frontally through pyramidal regions until it reaches area 8 (see below). This dorsal trend is mirrored by an equivalent 'ventral trend' (Barbas and Pandya 1987, 1991; 'lateral' according to Goldberg 1985), originating in olfactory allocortex and progressing frontally through granular regions until it too reaches area 8 (Fig. 1.9).

Cingulate cortex can also be viewed as originating from approximately the same allocortical primordium as the dorsal (lateral, hippocampal) trend in frontal cortex. However, in many respects it is quite distinct and, indeed, is functionally and connectionally more akin to the ventral (medial, olfactory) trend in frontal cortex. The anterior cingulate, while being physically located in frontal cortex (defined as cortex anterior to the central sulcus), has none of the posterior sensory cortical connectivity which distinguishes the adjacent dorsal trend. Nonetheless, it terminates (as do the other two prefrontal trends) at the border of area 6, with connections to area 6 and thence to area 4; and we shall argue that it is at least homologous with the frontal cortex in its relations to motor control. Indeed, one can view the cingulate motor cortex as providing access to the motor system for a large number of limbic areas in much the same way that frontal cortex provides access for sensory cortical areas (Fig. 6.8). The posterior cingulate is located behind the central sulcus and, we shall argue, is homologous with the posterior sensory cortices, and particularly with the temporal cortex and archicortex. Nonetheless, its thalamic connectivity is more akin to that of frontal cortex. Thus, cingulate as a whole can be viewed as quite distinct from both frontal and posterior cortex.

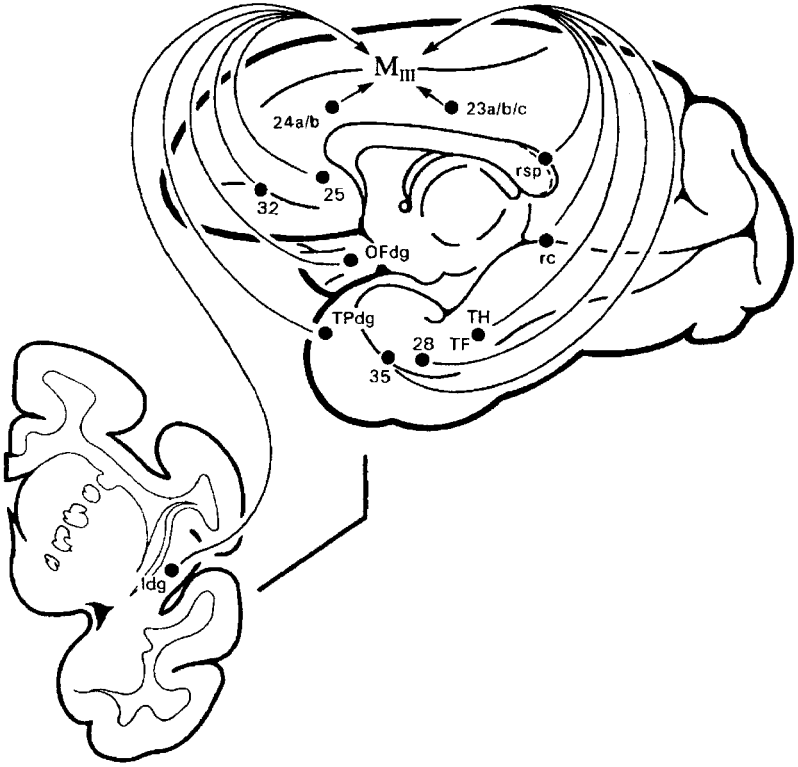


Fig. 6.8 Limbic inputs to the cingulate motor cortex. (M_{III}, area 24c) are widespread. Idg, insula (dysgranular); OFdg, orbital frontal cortex (dysgranular); rc, retrocalcarine cortex; rsp, retrosplenical cortex; TH/TF, temporal cortex areas TH/TF; TPdg, temporal pole (dsygranular). (From Morecraft and Van Hoesen 1998.)

The dorsal and ventral trends in frontal cortex are matched by, and primarily connected to, separate dorsal and ventral trends in each of the unimodal sensory cortices. Put crudely, the ventral trend in both frontal and posterior cortex deals with 'what' stimulus is (or may be) present; the dorsal trend deals with 'where' the stimulus is or may be (Gross 1994); and, we argue below, the cingulate deals with 'why' we should worry about the stimulus. Each of these areas is distinguished by its specific thalamic and sensory cortical connections. The dorsal trend receives largely external sensory input, while the ventral trend and then the cingulate have inputs which are progressively more weighted towards (but not exclusively devoted to) interoceptive/motivational inputs originating in the hypothalamus.

6.10 THE PREFRONTAL CORTEX

The therapeutic use of prefrontal lesions (prefrontal lobotomy, or the fibre-cutting operation known as 'leucotomy') has a long, rich, and alarming history. It was originally

introduced as a treatment for schizophrenia, for which, many thousands of gratuitous operations later, it was found to be of absolutely no use (Willett 1960). Nonetheless, it does alleviate symptoms of anxiety, obsession, and depression, symptoms upon which cingulate lesions also act (Marks 1969; Ström-Olsen and Carlisle 1971; Tan *et al*. 1971; Kelly *et al*. 1972; and see Chapter 11). However, with frontal lesions this alleviation comes at a price. In a review, Swayze (1995) noted the comments of the patients' families that the lesions 'destroyed the soul': the patients 'are described by the nurses and the doctors, over and over, as dull, apathetic, listless, without drive or initiative, flat, lethargic, placid and unconcerned, childlike, docile, needing pushing, passive, lacking in spontaneity, without aim or purpose, preoccupied and dependent' (Hoffman, cited by Swayze 1995, p. 507; see also Brown 1985). (Cingulate lesions have the advantage that they lack these major side-effects.)

Despite these practical problems, the clinical data apparently offer good evidence for some contribution of prefrontal cortex to anxiety; and, indeed, in the original formulation of the behavioural inhibition system, prior to the first edition of this book, we (Gray 1970b) attributed to the prefrontal cortex the role of cortical representation of this system. As we shall see, this view no longer fits either with the more recent version of the behavioural inhibition system theory as presented in Chapter 11 or with more recent views of the prefrontal cortex. These place considerable emphasis on functions such as working memory which appear largely intact in human beings treated with anxiolytic drugs. Thus, the 'loss of soul' described by Hoffman (above) may make an anxious patient more tractable without forcing us to conclude that frontal cortex is the source of the pathological anxiety.

As a first step to understanding the therapeutic effects of prefrontal lesions in human beings, it should be helpful to review the data on such lesions in other animals. Unfortunately, the literature in this field is fraught with multiple difficulties of interpretation. The first concerns the very meaning of the term 'prefrontal cortex'. It has been very difficult to determine precisely what is prefrontal cortex and into what subareas it can be divided. This has particularly been the case if one wished simultaneously to encompass findings from human subjects, monkeys, and rodents. The prefrontal cortex is itself normally defined by exclusion from the rest of frontal cortex, i.e. the cortical areas anterior to the central sulcus. It is 'the region anterior and mesial to those areas of frontal cortex having to do with motor functions (motor cortex and frontal eyefields) and speech (Broca's area)' (Benton 1991, p. 3). In addition, area 8, which overlaps the frontal eye fields, is occasionally included in prefrontal cortex. However, we shall treat the division between prefrontal and frontal cortex as arbitrary and suggest that there is, instead, a steady hierarchical progression from primary motor cortex to ever higher levels of processing, with no single categorical boundary dividing a set of 'motor' levels from a second 'non-motor' set (Fig. 1.9).

The monkey prefrontal cortex was early on divided into two functionally discrete regions (Brutkowski 1965). This had risen to between five and seven separate regions a decade and a half later (Rosenkilde 1979). These divisions were made almost purely on behavioural grounds. As a result, the boundaries of the regions and the functions attributed to them changed almost from experiment to experiment. The difficulties were still greater if the compass of interest extended beyond the monkey. Thus, in the rat, the prefrontal

cortex as a whole was not even clearly defined until Leonard (1969) established which regions receive projections from the mediodorsal nucleus of the thalamus similar to those which innervate the prefrontal cortex in the monkey. On the basis of these connections it was possible tentatively to identify two regions in the rat's rostral cortex as corresponding to monkey prefrontal cortex. Of the two major regions, one was thought to be roughly equivalent to monkey dorsolateral prefrontal cortex (Brutkowski 1965); in the rat this lies dorsomedially, extending ventrally into the medial walls of the hemispheres and caudally into the supracallosal region which was earlier treated as belonging to the cingulate cortex. The second was thought to be roughly equivalent to monkey orbital frontal cortex (Brutkowski 1965); in the rat it lies ventrolaterally, along the upper lip of the rhinal sulcus.

The earlier human data also relied heavily on behaviour-based anatomy, and, of course, generally with very unsystematic lesions to which to relate behaviour. What is more, at those points at which comparison between simian and human behaviour seemed most possible, there appeared to be major discrepancies between the cortical locations of different functions. Thus reversal learning was most markedly impaired in the monkey after damage to parts of the orbital frontal cortex (Deuel and Mishkin 1977; Rosenkilde 1979), whereas in man apparently similar deficits followed upon dorsolateral frontal damage (Milner 1963, 1964).

These definitional problems render much early work difficult to assess. Recent anatomical advances allow for a much clearer picture. Nonetheless, even recent lesion experiments cut across the newly discovered anatomical boundaries, while recent single-cell experiments have barely scratched the surface of anatomical field-specific analysis. Rather than attempt to resolve these problems retrospectively, in Appendix 3 we have employed three stratagems. The first is fairly straightforward and uncontentious. We borrowed wholesale from the conclusions of recent anatomical reviews of prefrontal cortex (while noting that these are heavily biased by analysis of non-human primates). The second is equally straightforward, but somewhat more contentious. We simply assert, as an initial working hypothesis, that the general parcellation and function of the rat and human homologues are essentially identical to those of primate prefrontal cortex, just as the general parcellation and function of the brain as a whole are essentially identical (Swanson 1995). The third is in some respects an act of faith: we proceed on the assumption that the frontal cortex, like posterior cortex, is organized in terms of a succession of levels of processing, each higher than the last, but of essentially the same kind.

There are two points to note in support of these assertions. First, recent advances in cytoarchitectonic analysis and in the investigation of functional organization have allowed quite detailed homologies to be constructed between human and macaque frontal cortex (compare Figures 5 and 6 in Petrides and Pandya 1994). Similar homologies have been suggested between primate and rat frontal cortex (see Kolb *et al.* 1994). In the latter case, functional analysis suggests that here also detailed homologies will ultimately be found with respect to primate frontal cortex, since 'there is a prefrontal region whose general functions appear to be generally similar across a wide variety of mammalian species' (Kolb *et al.* 1994, p. 678). The divisions based on this type of analysis are also being confirmed by more recent demonstrations that neurotransmitter binding-site densities 'change precisely at the cytoarchitectonic boundaries between different cortical areas'

(Gebhard *et al.* 1995, p. 509; see also Hof *et al.* 1995). This latter type of analysis opens up the possibility of quite detailed parcellation of rat cortex. Functional analysis with lesions in monkeys and positron emission tomography in human beings has shown that, at least for the comparison of areas 9 and 46 with area 8, the different species show the same double dissociation between self-ordered working memory and conditional association learning (Petrides and Pandya 1994).

Second, as a consequence of these new homologies, the previous apparent marked variations in rat, monkey, and human lesion data can be attributed to a failure to involve precisely the same cytoarchitectonic fields. However, even where the same field is involved in different species, some differences in behavioural effect are likely to arise, because the change in some common prefrontal function may be affected by variations in sensory input to, and cognitive capacity of, prefrontal cortex across species. Thus, 'the anatomical parcellation of subareas is significantly different across species, and it is likely that as the functional properties of the subareas are explored there will be significant interspecies differences in the details of prefrontal organization' (Kolb *et al.* 1994). Indeed, in man, there are significant differences in the precise parcellation even between individuals, which 'suggests that the individual differences among the classical maps of Brodman, von Economo and Kiskinas, and Sarkissov and others may have been due to normal variation among the brains they analysed. Such variations may underlie individual differences in the visuospatial and cognitive capacities subserved by these areas' (Rajkowska and Goldman-Rakic 1995b, p. 323; see also Rajkowska and Goldman-Rakic 1995a).

Variation in specific function does not necessarily imply, however, lack of common design. We would argue that there are consistent global functions subserved by the dorsal (medial) and ventral (lateral) trends in prefrontal cortex across species. However, the expansion of cortex which produces Broca's and Wernicke's areas in human brains must entail large differences in the precise organization of prefrontal cortex between people and monkeys, not to mention between people and rats. Nonetheless, we can see Broca's and Wernicke's areas as each functioning, within the linguistic domain, in a fashion similar to the non-linguistic functioning of their respective surrounding cortices. Our analysis, therefore, has concentrated on the non-human primate data. These are more numerous than the equivalent human and rat data. Given the enormous expansion of the human frontal cortex relative to other primates, the non-human primate can, in a sense, be viewed as representing an approximate half-way point between the rat and man.

The view we present of prefrontal cortex (Appendix 3) amalgamates two apparently separate types of theory (see, for example, Passingham 1993, pp. 208–9). The first is the theory that prefrontal cortex is the apex of the motor system. On this view, prefrontal cortex is the highest, executive, level of a hierarchically organized motor programming network, and 'is concerned with the process by which new decisions are taken as to what to do' (Passingham 1993, p. 209). The second theory (e.g. Goldman-Rakic 1992) holds that prefrontal cortex subserves working memory (or, better, the more general 'active memory'; see Fuster 1995). On this view, information from many cortical areas is transferred to prefrontal cortex for use in current processing, and prefrontal cortex is the highest level of a hierarchically organized perceptual analysis network (Fig. 6.9).

While these proposals sound quite distinct, they are not incompatible. Indeed, if we accept MacKay's (1987) view that perception and action must often involve the same neural circuitry, amalgamation of these points of view is not only easy but almost tautologous (see also Fuster 1990, on the 'perception–action cycle').

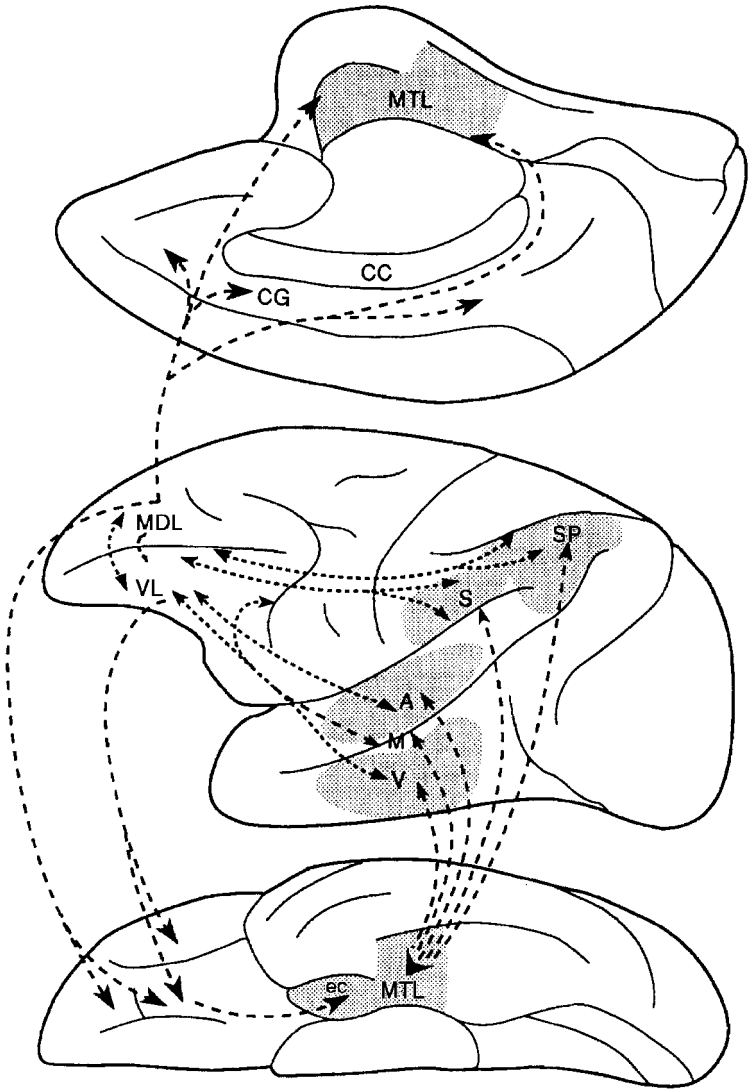


Fig. 6.9 Frontal–sensory cortical connections which may subserve working memory. The reciprocal connections could subserve the transfer of information from the sensory to the frontal cortex or could allow frontal cortex to refresh and keep active circuits encoding particular types of information in sensory cortex. Final processing stations for sensory information are indicated by: A, auditory; M, multimodal; S, somatosensory; SP, visuo–spatial; V, object vision. MDL, mid-dorsolateral frontal cortex; MTL, medial temporal lobe; VL, ventrolateral frontal cortex. (From Petrides 1994.)

6.11 THE CINGULATE CORTEX

The *gyrus cinguli*, together with the hypothalamus and hippocampus, was proposed by Papez 'as representing theoretically the anatomic basis of the emotions' (Papez 1937, in Arnold 1968, p. 302). It receives major input from the mammillary bodies of the hypothalamus via the anterior thalamic nuclei and so 'may be looked on as the receptive region for the experiencing of emotion as a result of impulses coming from the hypothalamic region, in the same way as the area striata is considered the receptive cortex for photic excitations coming from the retina' (Papez 1937, in Arnold 1968, pp. 305–6). Given the input from the anterior thalamus, cingulate cortex should be viewed as being more similar to frontal than to simple sensory receptive cortex, since major input from the anterior and mediodorsal thalamus is one of the features which conventionally distinguishes frontal from posterior cortex. Cingulate cortex can, in particular, be related to the ventral trend of prefrontal cortex to which the thalamus relays hypothalamic information. A relation to prefrontal cortex is particularly obvious with regard to anterior aspects of the cingulate, which are contiguous with motor, premotor, and frontal cortex proper. However, we also argue (Appendix 3) that emotional perception has a much tighter relationship to action than to perception of the external world. As a result, despite its links with frontal and particularly motor cortex, we can still view cingulate as 'the receptive region for the experiencing of emotion'. This latter view is also consistent with the contiguity of posterior aspects of cingulate cortex to somatosensory cortex.

The cingulate cortex can be regarded as consisting of two fundamentally distinct (but reciprocally connected) modules. Cingulate cortex posterior to the central sulcus is related to (but not part of) posterior cortex (with its thalamic input playing something of the role of, for example, the geniculate and collicular input to striate cortex). Cingulate cortex anterior to the central sulcus is related to (but not part of) frontal (motor/premotor) cortex (with its thalamic input playing a role equivalent to that of the anterior and dorsomedial thalamic input to frontal cortex). However, the thalamic nuclei which provide input to the posterior cingulate include ones which project to frontal cortex; and, given its input from pain pathways, anterior cingulate acts in part as primary sensory cortex. So, both anterior and posterior cingulate combine, in their different ways, features from both frontal 'action' systems and posterior 'perception' systems. In this sense cingulate cortex is quite distinct from both frontal and posterior cortex, and we shall treat it (together with the septo-hippocampal system) as a third major 'mesial' division of the cortical mantle. On this view, posterior and anterior cingulate cortex are differentiated, not by relations to posterior and anterior cortex respectively, but, as with frontal and posterior cortex, by the division of mesial cortex into parallel 'dorsal' and 'ventral' trends.

Thus, while the cingulate cortex as a whole receives input, relayed by the thalamus, from the mammillary bodies, and while posterior cingulate receives direct input from the lateral hypothalamus and the ventral tegmental area, there are many reasons why we can treat the cingulate neither as a homogeneous area nor solely as primary receiving cortex for hypothalamic information (Vogt 1985). We must divide cingulate cortex into a number of discrete areas, exactly as with frontal cortex, on the basis of parallel 'dorsal' and 'ventral' trends, each subdivided into architectonically progressive levels (Fig. 1.9).

6.12 OVERVIEW OF PREFRONTAL AND CINGULATE CORTEX

The prefrontal cortex is the seat of gaze control, working memory, and yet higher functions of the motor hierarchy associated with the pre-planning of non-innate movements. For working memory to occur, long-term memories and current percepts originating in primary and association sensory areas (including, for example, spatial information stored in the parietal cortex) must be loaded, by direct connections from these areas, into prefrontal cortex in the region of the principal sulcus. It seems necessary, and there is some evidence for this, that there should be topographic organization of the prefrontal cortex, allowing different types of information to be kept separate. Such topographic mapping is likely also to be important for any planning function. In order to mediate this function, activity arising in a specific area of frontal cortex in the absence of specific environmental stimuli would initiate the activation of the appropriate part of a specific posterior cortical system (hence providing a basis for imagery).

Working memory information is an integral part of the distributed (and presumably recursive) activity in the prefrontal cortex which controls ongoing motor acts. As noted above, it is best in this connection to blur the distinction between prefrontal and motor areas. As we progress from the most differentiated areas of frontal cortex to the least (e.g. 4–6–8–46–10–14–25), we appear to deal with progressively more anticipatory aspects of motor control (motor–premotor–visual attention–working memory–planning). Working memory appears to acquire its 'working' aspects from the recursive interconnections of the principal sulcus with other frontal areas (more motor in the case of the more differentiated areas of posterior frontal cortex, more planning in the case of the less differentiated areas of anterior and ventral prefrontal cortex), and its 'memory' aspects from the recursive interconnections of the principal sulcus with posterior cortical areas. On this view, working memory generally would function in an analogous fashion to the 'articulatory loop' which supports the subvocal rehearsal of verbal material (Baddeley 1986, Chapter 5). Although we lack a specific psychological terminology for the processes involved in each of the less-differentiated levels of prefrontal cortex, as we move from gaze control to working memory to these less-differentiated levels it appears likely that the fundamental architecture is the same. The links to the same architectonic level of posterior cortex provide the primary sensory information required, and the links to more and less architectonically differentiated levels of frontal cortex provide coordination between more and less anticipatory levels of pre-motor control.

Our admittedly speculative and incomplete formulation tends to amalgamate many current views of prefrontal function (see Grafman 1995, for a comparative review), and does so by ascribing each of those views in essence to a different level of frontal cortex; or, as in the case of working memory, by asserting that both of two apparently distinct views (memory versus planning) are simultaneously correct. Grafman complains that all of these current views are weak (in which case, so is ours) because they deny frontal circuits 'any special cognitive status/role other than that they manipulate knowledge that is primarily stored in posterior regions of the cerebral cortex' (Grafman 1995, p. 344). To this, we unrepentantly plead 'guilty as charged'. In our view perception, long-term

memory, and active memory for any particular cognitive entity all reside in the same cell assembly, which can be activated by external input, contain strengthened synapses, or be activated by frontal cortex, respectively, to fulfil these different functions. However, it is implicit in our ascription of 'planning' and 'intention' to the frontal cortex that this region should in addition contain its own cognitive representations. Thus, while we agree with Grafman (e.g. Grafman 1995, p. 360) that frontal cortex may, in essence, store plans, scripts, schemata, themes, or mental models, it seems likely that it does so through its capacity to order or relate items, these themselves being stored elsewhere. Thus the *knowledge* embodied in a 'plan' would be distributed outside the frontal cortex, but without information stored in the frontal cortex it could not be integrated into a *plan*. Consistent with this formulation, the information encoded by dorsolateral frontal cortical neurons during a delay appears in at least some cases to be a representation of the anticipated upcoming goal rather than of the prior stimulus which signalled that goal (Watanabe 1996).

The loss of behavioural spontaneity after lesions to the dorsal trend of frontal cortex, on this view, is due to the disconnection from the motor system of the capacity to predict in advance critical temporal or spatial aspects of the world. Working memory, in this case, can be thought of, not so much as storing information *per se*, but as predicting where an upcoming, delayed, response must be made (e.g. Funahashi *et al.* 1993). Dorsal prefrontal lesions, thus, remove the capacity to initiate action under conditions in which a degree of anticipation is required to achieve particular goals (the 'where' function). They leave intact, however, posterior cortical mechanisms which allow fairly direct reactions to stimuli. Even in the supplementary motor area (area 6, medial premotor cortex), the primary function appears to be the selection of responses on the basis of internal cues, with the result that the provision of external cues, or instructions, can eliminate the problem with motor control resulting from lesions to this region (Chen *et al.* 1995; Thaler *et al.* 1995; see also Viallet *et al.* 1995). By contrast, the release of behaviour by lesions to the ventral trend of frontal cortex would, on this view, be due to the disconnection from the motor system (and from the dorsal trend of prefrontal cortex) of the capacity to predict in advance emotionally significant aspects of the world (the 'what' function). A dysfunction of this kind is likely to be of greatest importance in relation to affectively negative aspects of the world. With a properly functioning ventral trend, prediction of upcoming aversive events can inhibit the more immediate responsive initiation of action by posterior cortex, as well as the less immediate initiation of action by the dorsal trend of prefrontal cortex. Note that, in this respect the function of the ventral trend of prefrontal cortex conforms to that proposed for the behavioural inhibition system, and hence potentially to a role in anxiety. However, we qualify this idea below.

On this view, it is the necessity of predicting events of significance by the ventral trend of prefrontal cortex which requires it to receive input from olfactory cortex. This is one of the phylogenetically oldest sources of warning of upcoming events; indeed, the simple detection of an odour, with no additional analysis, can provide a warning. Likewise, it is the affective loading of predictions by the ventral trend of prefrontal cortex which requires it to receive thalamic information similar to that received by cingulate, and to receive input from the cingulate cortex itself. But if this is the role of the ventral trend

of frontal cortex, what is the role of cingulate cortex? Given what we have just said, why is cingulate cortex necessary at all? To answer this question, we must remember the tendency of the brain to produce parallel structures for related functions (e.g. the dorsal and ventral trends for 'where' and 'what', and the multiple topographic representations of visual space for the extraction of form, movement, and colour). The data we have considered above link prefrontal cortex with the anticipatory control of motor programs, the steps of which are controlled largely by external events and by memories of such events coded by posterior cortex. In a complementary fashion, they (and other data considered in Appendix 3) link cingulate cortex with the anticipatory control of motor programs, the steps of which are controlled largely by internal events coded as somatosensory input (e.g. nociception) or as hypothalamic input.

We can now see why, on a small scale, cingulate appears to mirror prefrontal cortex. It undertakes the same general tasks but with largely 'innate' as opposed to 'acquired' motor plans (this does not mean, however, that the stimuli which elicit the plans cannot be learned). Anterior cingulate presumably adds greater range and anticipatory character to the initiation of, for example, the simple (UCS–UCR) active avoidance reactions which can be generated via the amygdala. (In particular, it is likely to operate in an anticipatory fashion, for example, when there is no explicit UCS to elicit the UCR.) Similarly, posterior cingulate may add greater range and anticipatory character to the inhibition of ongoing behaviour and the elicitation of risk analysis which can be generated via the hippocampus. Obsessions, then, can be viewed as the activation (pathological or non-pathological) of largely innate affective motor plans within the cingulate cortex (which could involve the placing of an item into the working memory areas of prefrontal cortex). At least one source of *pathological* obsessions could then be a failure to provide to the cingulate a 'task completed' feedback signal from the motor system analogous to that which, according to Goldman-Rakic (1992), terminates the delay firing of principal sulcus neurons (Fig. 6.10). As with the ventral trend in frontal cortex, then, posterior cingulate has functions which conform to our description of the behavioural inhibition system and could, then, be involved in anxiety; again we qualify this inference to some extent below.

Given their parallel functions, it is not surprising that the operation of cingulate cortex is likely to be modified by output from the prefrontal cortex (Pandya *et al.* 1971). Reciprocal interconnections between these areas would be necessary to coordinate their distinct roles in the control of ongoing behaviour. There are also close connections between the prefrontal cortex and the septo-hippocampal system. In monkeys there are prefrontal projections to the entorhinal cortex (Van Hoesen *et al.* 1972, 1975; Van Hoesen and Pandya 1975), to the hippocampus (Leichnetz and Astruc 1976), to the septal area (Johnson *et al.* 1968; Tanaka and Goldman 1976) and to the mammillary bodies (Jacobson *et al.* 1978). All of these components of the septo-hippocampal system can influence the cingulate cortex fairly directly. Similarly, all areas of cingulate cortex project to entorhinal cortex (Suzuki and Amaral 1994) and hence can influence the prefrontal cortex indirectly. Note, however, that while there is good evidence for homologous functions of rat prefrontal and primate prefrontal cortex, there is, as yet, no evidence that the descending connections just described for the monkey are present in the rat. Thus, while the possible link between prefrontal cortex and anxiety which is

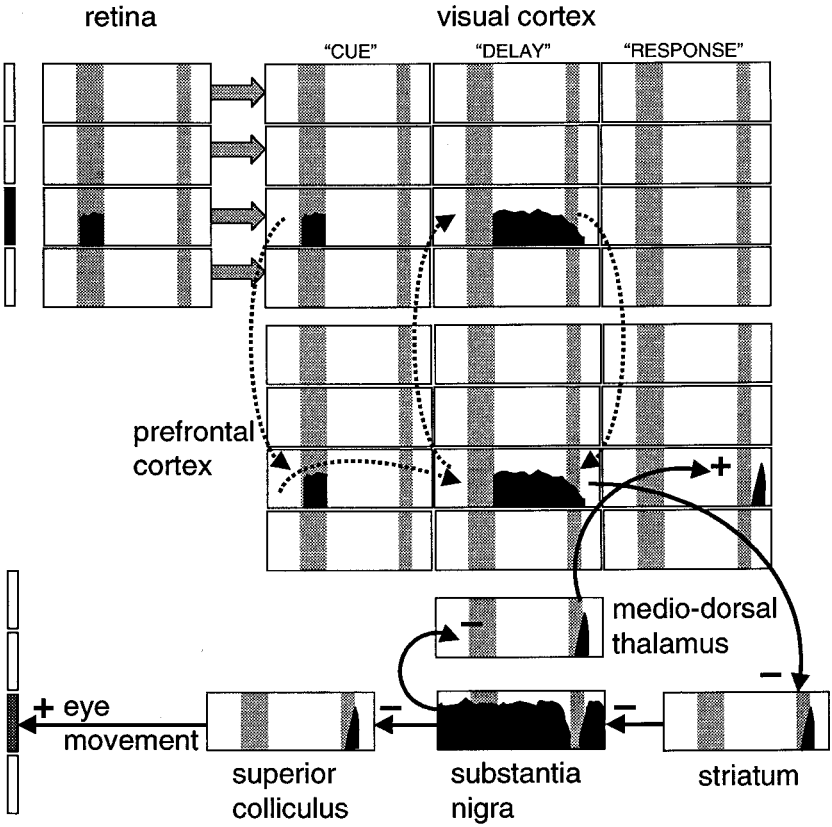


Fig. 6.10 A schematic diagram showing the possible neural architecture underlying responses in a simple working memory task in which the monkey must fixate the centre of a screen, remember the position of a dot during a delay, and then on presentation of an end-of-delay signal move its eyes to the position at which the dot was previously displayed. Presentation of a dot at a specific point in visual space (filled bar at top left) activates the retina topographically. The retinal input to visual cortex produces topographically mapped activity in two classes of cell 'cue' and 'delay'. Because of the significance of this input for the upcoming motor response, activity in the cue cell produces topographically mapped activity in prefrontal cortex cue cells and hence delay-related activity in prefrontal delay cells which interact with visual cortex delay cells to maintain the visual information during the delay. On presentation of the end-of-delay signal there is a release of inhibition of motor circuits which allows the animal to move its eye to the position maintained by the activity in the prefrontal-visual cortical loop. There is a feedback signal, via the medio-dorsal thalamus, which resets the working memory system. (Adapted from Goldman-Rakic 1992.)

indicated here is likely to be phylogenetically old, the role of cortical structures in anxiety appears to have greatly expanded at the primate level. We shall argue later, however, that this expansion is purely quantitative and does not change the general qualitative principles of operation of the networks controlling anxiety.

6.13 PREFRONTAL AND CINGULATE CORTEX AND ANXIETY

We have, then, suggested a role for the posterior cingulate cortex in the generation of what could be called innate anxiety plans, and for the prefrontal cortex in the working memory and motor programming functions required for acquired anxiety plans. These proposals raise two questions. First, can either the cingulate cortex or the prefrontal cortex be seen as the sole seat of anxiety? Second, are they likely to be principal sites of action of anxiolytic drugs?

Certainly, there are superficial parallels between the anxiolytic/septo-hippocampal syndrome and the effects of prefrontal lesions. In rats (which have been the main species used to define the anxiolytic/septo-hippocampal syndrome), prefrontal lesions impair water maze performance, radial-arm maze performance, Lashley III maze performance, delayed response, delayed alternation, spontaneous alternation, and delayed non-matching to sample, but not matching to sample. They also impair discrimination reversal, performance on a DRL schedule, and extinction. Consistent with the idea that changes in the delayed non-matching task reflect a loss of working memory and that different areas code for different types of working memory, there appears to be some differentiation between deficits in visual working memory produced by more medial prefrontal lesions and deficits in olfactory working memory produced by more orbital regions. (For all of the above see Kolb 1984; Kolb *et al.* 1994.) Further, frontal deficits appear to be selective for working as opposed to reference memory and are strongly affected by interference (Granon *et al.* 1994). In a direct comparison of septal and prelimbic lesions, Brito and Brito (1990) found that both had similar effects on open field behaviour, activity, response patterning, and go–no-go discrimination; and that neither lesion affected visual discrimination, step-through passive avoidance or active avoidance. However, these authors also found that septal but not prelimbic lesions increased eating in a novel environment, impaired performance of learned alternation (patterning), impaired delayed non-matching to position, improved olfactory discrimination, improved two way avoidance, and increased shocks taken in an approach–avoidance conflict. In contrast to this report of largely unimpaired aversive conditioning in prefrontal rats, Frysztak and Neafsey (1994; see also the single-cell recordings in Maxwell *et al.* 1994) reported a partial impairment in a measure of a conditioned emotional response, namely, heart rate conditioning (which is normally not impaired by hippocampal lesions, but is seriously impaired by amygdalar lesions); and Kolb's (1984) review suggests that one-way and two-way active avoidance, but not passive avoidance, are impaired by medial frontal lesions.

In both human beings and rats, there appear to be a number of other dissociations, some of which may be related. For example, item recognition and order recognition are both impaired by hippocampal lesions, but only order recognition is impaired by prefrontal lesions (Kesner *et al.* 1994b; for a lack of effect in monkeys see Colombo *et al.* 1993). Successive odour discrimination is impaired by frontal but not hippocampal lesions (Eichenbaum *et al.* 1980; Cohen and Eichenbaum 1993), as is the temporal order function in paired associate learning (Kesner 1993). It also appears that in at least some cases of conditional and non-conditional matching to sample, prefrontal lesions produce delay-independent and hippocampal lesions delay-dependent impairments (Winocur

1992; Chao and Knight 1995; see also Pigott and Milner 1994). However, it should be noted that none of these studies used bias-free measures of performance or fitted exponential decay curves (see Chapter 7, Section 7.4); and, in the study by Chao and Knight (1995), there was a major problem with ceiling effects. Hippocampal but not frontal lesions impair recall of incidentally learned locations (Smith and Milner 1984). Prefrontal but not hippocampal lesions produce impairments on the Wisconsin card sorting test, while the reverse is found with face recognition, recall of spatial location, and 'on various perceptual and memory tests that have proved difficult for patients with left or right temporal lobe lesion' (Milner 1995, p. 69).

There are thus positive and negative discrepancies between the effects of prefrontal and septo-hippocampal lesions and, despite the very limited data available, it appears to be the septo-hippocampal rather than the frontal lesions which have effects like anxiolytic drugs (note particularly the cases of approach–avoidance conflict and two-way avoidance). We have already commented on the fact that prefrontal and cingulate cortex lesions are used for treating anxiety specifically in those patients who do *not* respond to anxiolytic drugs. There must, therefore, be structures other than frontal and cingulate cortex which contribute to anxiety and on which the drugs normally act; and, as we have already argued, the septo-hippocampal system seems the principal candidate. Equally, the fact that both cingulate and prefrontal lesions are used to treat anxiety implies that neither region can be the sole seat of even that residual anxiety which is not affected by anxiolytic drugs (see also Sachdev and Hay 1996, for problems with identifying the appropriate size and site of such lesions). If both can contribute to anxiety, then we need to consider what relation they have to each other and to other structures identified as playing a part in the control of anxiety. This will be discussed in Chapter 11.

However, before moving on, we must consider one additional feature of anxiety in human beings which is tightly linked to the functions we have described for prefrontal and cingulate cortices: verbal behaviour. In several studies of the effects of prefrontal damage in man it has been observed that there is a dissociation between, on the one hand, the patient's ability (intact) to describe the requirements or contingencies of the task they are set and, on the other, their ability (impaired) to control their behaviour in accordance with this description (Rosenkilde 1979). This is similar to the dissociation observed with cingulate damage between reportable pain detection (increased) and affective value (decreased). Note that this dissociation is opposite in sign to that of the amnesic syndrome consequent upon damage to the hippocampus. In the latter, memory is sometimes behaviourally intact in cases where the patients cannot describe the contingencies which control their behaviour. Intact verbal report of contingencies with impaired behavioural control in patients with prefrontal lesion has been reported in a go–no-go discrimination (Drewe 1975a,b; Toczek 1960) and on the Wisconsin card-sorting test (Milner 1963, 1964). The fact that the patients were able to describe correctly the contingencies of reinforcement operating in these experiments, yet could not use this knowledge to control their behaviour, suggests that in human beings the prefrontal cortex serves as a way station by which language systems normally *can* contribute to the inhibition of incorrect behaviour.

This conclusion implies that frontal cortex could provide an extra layer of inhibitory control, dependent on highly processed cortical information, superimposed on psycho-

logically simpler subcortical and archicortical control. One specific suggestion, which will be important for the development of our theory, is that the performance of prefrontal lesioned patients 'is best explained in terms of an inability to see or resolve a goal–subgoal conflict. This interpretation is compatible with several existing accounts of frontal lobe dysfunction that postulate a failure of inhibition of a prepotent response to explain poor performance on the Wisconsin Card Sorting task, the Stroop task, the Antisaccade task, the A-Not-B task, and the Delayed Alternation task' (Goel and Grafman 1995). We shall argue accordingly that the similarities and differences between frontal and hippocampal damage arise because the former is concerned with resolving goal–subgoal conflicts. This includes ordering goals in the correct sequence (Funahashi *et al.* 1993; Sirigua *et al.* 1995), essentially by inhibiting actions which have been completed (Kolb 1984) as well as those which are not yet required. By contrast, the hippocampus is concerned with resolving goal–goal conflicts, and, at the next level down, the motor system is concerned with resolving goal–motor command conflicts of the sort that arise in, for example, mirror drawing.

Such a development would make good evolutionary sense. The hierarchical defensive systems which we described earlier have at their subcortical apex the hippocampal formation, whose function we characterized as that of correcting motor programs when things go wrong. This is the sort of problem on which, typically, human beings bring to bear their full powers of verbal analysis. Thus it would not be surprising if, in the human brain, some sort of liaison were established between the language systems of the temporal and frontal cortex (Lenneberg 1967) and the septo-hippocampal system, so as to arrange, as it were, mutual cooperation. This could take place by way of septo-hippocampal activation of appropriate language circuits, by way of descending cortical control over the septo-hippocampal system, or both. The fact that prefrontal lesions leave verbal descriptions of behavioural contingencies intact, while the behaviour itself fails, suggests that the prefrontal cortex lies athwart the second of these two routes.

The notion that prefrontal lesions disconnect the septo-hippocampal system from the language system is supported by the studies of Homskaya (1964), who showed that in normal subjects or patients with other brain lesions, the galvanic skin response and vasomotor orienting response (Sokolov 1960) appear and disappear rapidly in response to verbal instructions (e.g. the patient is told that a stimulus will, or will no longer, be followed by pain). But these responses to *verbal* stimuli are absent in patients with prefrontal damage. Conditioning and extinction of the galvanic skin response and vasomotor response to *non-verbal* stimuli, while not entirely normal, did take place in prefrontal lesioned patients. Thus the disturbance produced by prefrontal lesions was relatively selective to verbal control of orienting responses.

These observations suggest an explanation of the powerful effects of prefrontal lesions on human anxiety. For this we require the assumption that the circuit connecting, by way of the prefrontal cortex, language mechanisms and the septo-hippocampal system permits activation of the latter by verbally coded description of threats. These pathways would constitute an alternative alarm bell acting instead of, or in tandem with, the phylogenetically older systems which activate the amygdala and septo-hippocampal system. If this hypothesis is correct, one would expect the 'cortical alarm bell' to be relatively unimportant in the anxiety elicited by innate anxiety stimuli or innate fear stimuli

(Seligman 1971; and see Chapter 11, pp. 277–8) such as snakes or spiders. Conversely, the cortical system would play a relatively large role in anxious, phobic, or compulsive behaviour that depends heavily on such processes as semantic generalization (Razran 1971), verbal self-description of potential threats, or the perception of situations as threatening only by virtue of their verbal description (e.g. the possibility that one will fail to meet a self-imposed deadline at work). These considerations are *consistent* with—but the argument is too tenuous to permit a stronger term—the type of patient who has been reported to benefit from psychosurgery: extreme neurotic introverts displaying obsessional symptoms, widely generalized phobic behaviour, or anxiety states (Marks *et al.* 1966; Tan *et al.* 1971; Powell 1979).

This view is also consistent with the impressive success of cognitive restructuring and related psychological methods in treating anxiety disorders (Dalgleish and Power, 1999; see Chapter 13). Whether the disorder *originates* in the prefrontal area, language area, or lower parts of the defence system, it is clear that appropriate alterations to the language system would be expected to influence the prefrontal cortex and through it the cingulate, hippocampus, and amygdala. It should also be noted that other high-order cognitive abstractions, akin to language and similarly dependent on cortical processing, could well take the same route. This also fits with the general pattern of defence system dominance which we have observed, where lower levels are 'primed' and will also control action unless a higher level indicates that it is taking over control.

The fact that we have adopted the hypothesis of a descending cortical alarm mechanism travelling by way of the prefrontal cortex does not eliminate the possibility that there is also traffic in the other, corticopetal direction. For example, the subiculum (output station of the septo-hippocampal system) could influence much of the prefrontal cortex via its projection to the cingulate cortex (Meibach and Siegel 1977) as well as via its more limited direct projections to the temporal lobe.

If these arguments are correct, the role of cortical structures in anxiety has become much greater in man; and common-sense considerations make this conclusion almost a truism. However, the role suggested by Goldman-Rakic for the prefrontal cortex is much broader than the mainly linguistic one we have just been considering (see, for example, Fig. 6.10). If working memory and its more sophisticated relatives are, as seems likely, as much part and parcel of threat processing as they are of motor control, then we would expect the prefrontal cortex to play a role in animals comparable to that which it plays in man, with a quantitative rather than qualitative increase in its influence as its relative size increases. Indeed, it has been suggested that our linguistic capacities are simply an evolutionary elaboration of the pre-existing properties of frontal networks in other primates (which are themselves unusually elaborate; Aboitiz 1995).

6.14 CONCLUSIONS

We have now ascended the hierarchical defence system to its top. We have gone from the periaqueductal grey, which controls the simplest, undirected, escape responses, to cortical systems which control the most complex and anticipatory avoidance responses. These levels can be thought of, in many respects, as varying in terms of the defensive

distance with which they deal. At the higher levels of the system we have also outlined a distinction in terms of defensive direction, with, for example, amygdala and anterior cingulate cortex being largely concerned with active avoidance in comparison to the septo-hippocampal system and posterior cingulate corex, largely concerned with passive avoidance (and its related risk assessment).

Our review has shown that the match between the behavioural effects of lesions of a structure and those of injections of anxiolytic drugs (as outlined in the preceding chapter) is greatest in the case of the septo-hippocampal system. To some extent this appears to be because many of the structures on which anxiolytics have their direct actions (locus coeruleus, raphe nuclei, supramammillary nucleus, amygdala) have connections with the septo-hippocampal system, each providing a separate component of the overall anxiolytic/septo-hippocampal syndrome. One important exception to this generalization, however, is that the 'increase arousal' output of the behavioural inhibition system is clearly not mediated by the septo-hippocampal system, but rather by the amygdala.

Our overall conclusion (with only minor caveats), then, is that the bulk of the key effects of anxiolytic drugs are achieved directly or indirectly by disturbance of the functions of the septo-hippocampal system. Recall, however, that we found in Chapter 5 reason to question whether anxiolytic action (which itself, as we have seen in the previous section, is a guide to only some types of anxiety) is directed to a single structure. The largely correlational argument we have mounted in favour of the septo-hippocampal system as the core of the behavioural inhibition system and of the control of anxiety needs, therefore, to be backed up with a detailed review of the functions of the septo-hippocampal system, as well as direct evidence for its mediation of the behavioural effects of anxiolytic drugs. We leave until after we have discussed those functions in detail the question of how far the septo-hippocampal system is directly or indirectly involved in the various anxiety disorders as categorized by DSM-IIIR and DSM-IV (American Psychiatric Association 1987, 1994; see Chapter 11). This discussion will be the purpose of the next part of the book—an analysis of the septo-hippocampal system.

7 Hippocampal place fields

This and the following three chapters are devoted to an analysis of the functions of the septo-hippocampal system. One of the threads that continually criss-crosses through the intense and variegated debates centred on this issue concerns the degree to which the hippocampal system deals specifically or exclusively with spatial cognition. Without special experimental designs, it is usually difficult in any particular paradigm to separate its spatial and non-spatial elements. To do justice to this aspect of the debate, therefore, requires a kind of double vision, in which one considers almost every experimental result from both spatial and non-spatial points of view. To spare the reader this burden, we have chosen to deal first, in this chapter, with the nature of the environmental features in response to which hippocampal cells preferentially fire (since data of this kind provide the most immediate and compelling evidence for the spatial view). The chief conclusion we shall draw is that hippocampal cells represent, not spatial location *per se* nor exclusively, but 'available goals' (a concept which will become clearer as the chapter proceeds), be these defined spatially (as, given the nature of the environment in which animals must survive, is predominantly the case) or non-spatially.

Undoubtedly one of the most important contributions to an understanding of hippocampal function has come from the theory that this is concerned with spatial mapping (O'Keefe and Nadel 1978). A key plank in the construction of this theory lies in the so-called 'place fields' of hippocampal cells. Indeed, these provided the initial impetus for the theory, and their properties and the neural basis of their control are still being intensively studied more than 20 years after their discovery. Any essentially non-spatial theory (such as ours) must provide an account of them. The present chapter accordingly presents a brief survey of the data from experiments in which recordings of the firing patterns of single hippocampal neurons have been related to behaviour in spatial tasks. A more detailed review can be found in Appendix 6. Strictly speaking, the spatial theory of hippocampal function is a theory of spatial memory, since it holds that spatial maps are not only formed by the hippocampus but stored in that structure (Nadel 1991). We shall therefore take a further look at the more general aspects of the theory in the next chapter (Section 8.13), when we deal with memory.

Before presenting the observations that relate specifically to spatial cognition, we start this chapter with a preview of data derived from single-unit recordings in relatively simple experimental situations, in which the spatial element is minimal. These data are more extensively discussed in Chapter 10, in which they figure, as it were, in their own right. Here, they provide only a background before we go on to the more complex tasks that have provided the most striking evidence in favour of the spatial mapping theory. Since we present this background in a highly condensed form, the reader may prefer to read the appropriate section (10.3) of Chapter 10 first.

7.1 SINGLE-UNIT RESPONSES: NON-SPATIAL TASKS

With simple tasks (habituation to a novel stimulus; simple CS–UCS pairing; and simple simultaneous discriminations), a number of workers, most notably Olga Vinogradova (1975), have attempted to determine not only the behavioural correlates of hippocampal cells but also the relation between the firing patterns of cells at different points in the hippocampal formation and in its immediate subcortical and cortical neighbours. Their results provide us with a foundation for understanding the more complex firing patterns in more complex tasks.

Overall, it appears that the medial septum provides an input which, because of its short latency and multimodality, is likely to be relatively uninformative except as to the fact that there may be a need to cope with a situation (i.e. a goal is present and subcortical, e.g. orienting, response mechanisms, are primed). By analogy with the subcortical visual input to the amygdala (Chapter 6), this is a 'quick and dirty' pathway. In the absence of a suitable pre-programmed reaction (signalled via the entorhinal cortex), output from the hippocampus will elicit exploratory behaviour (cf. simple avoidance as elicited from the amygdala). After a number of occurrences of a stimulus, neocortical areas build up a model of the stimulus and of any required responses. These goals are passed to the hippocampus by the entorhinal cortex, with the final step of model building (or at least plasticity) occurring in the dentate gyrus. An important point is that separate models of the stimulus must be passed simultaneously for all current potential goals if inappropriate exploration is not to be elicited.

In the absence of modulatory aminergic input, the coincidence of septal and entorhinal input eliminates hippocampal output (but if *either* input occurs alone, then an output is generated). In the presence of aminergic input, weak inputs are suppressed by presynaptic inhibition (matching the inhibition of spontaneous single-cell activity seen in paired pulse paradigms), while strong inputs are facilitated by postsynaptic increases in excitability (matching the potentiation of population spikes seen in heterosynaptic paired pulse paradigms). This results in potentiation continuing to progress into the hippocampus, i.e. increasing cellular responses. The results of Foster *et al.* (1987) indicate that the activating effects of reinforcement and local habituation-like changes resulting from predictable sequences can summate with each other. In these experiments, there was direct evidence for what could be called a 'logical gate' (Section 10.3) only between the dentate on the one hand and CA1 and CA3 on the other. However, there appears to be a series of logical conditions which can result in augmented activity in dentate only, or dentate and CA3 only, or dentate and CA3 and CA1 only, or through the entire trisynaptic circuit, or, finally, in the targets of hippocampal output. No doubt more complex combinations are also possible.

Note that Vinogradova's (1975) data indicate that long-term potentiation (see Chapter 10, Section 10.3; Appendix 6, Section A6.2.1.8) of dentate input *cancels* rather than enhances hippocampal responses. She has shown (Vinogradova, personal communication) furthermore that, if long-term potentiation is artificially induced by perforant path stimulation, the hippocampus proper becomes totally unresponsive to natural stimuli which previously elicited a response (see also Miller *et al.* 1995). These results, taken together with the data from Thompson's group (Appendix 6, Section A6.2.2) showing

an extra-hippocampal locus for the plasticity underlying eyeblink conditioning (consistent with the effects of hippocampal lesions; Appendix 8), imply that the hippocampus is not the site at which a model of any stimulus, response, or goal is stored. If it were, experience would enhance rather than eliminate hippocampal responses, and removal of the hippocampus would eliminate rather than leave intact goal-directed behaviour. Rather, the hippocampus appears to receive a *copy* of the output from a model constructed elsewhere in the brain. It can then serve as a location at which the requirement for, for example, exploration is indicated by any mismatch between actual and modelled input. The extra-hippocampal location of the representation of the goal and the receipt by the hippocampus of an efference copy of its activation are both well illustrated by the data on eyeblink conditioning. During simple conditioning tasks, the firing of hippocampal cells closely tracks the nictitating membrane response (rather than the stimuli which give rise to that response; Berger and Thompson 1978a,b). Nonetheless, hippocampal lesions do not affect in any way the production of the nictitating response that has been so modelled (Appendix 6, Section A6.2.2). Presumably, a pattern of firing similar to that observed in hippocampal cells must exist in the effector circuits themselves, so as to produce the observed muscular response. It seems parsimonious to assume, therefore, that these circuits send on to the hippocampus information about their own activation, i.e. an efference copy.

Note also that, matching the parallel septal and entorhinal inputs to all levels of hippocampus proper (demonstrated anatomically), there is good evidence that quite specific entorhinal information can frequently be passed directly to area CA1 and that the same information, once filtered by the dentate gyrus and the septal inputs, can appear effectively degraded within area CA3 before being passed on 'to meet itself' in CA1. It is, of course, not 'the same information' and the entire purpose of the linear organization of hippocampal fields may be to provide a hierarchical series of logical gates, with output possible from each level of the system (see Fig. 10.3).

The experiments we consider in this section have used both simple stimuli and simple learning paradigms in rats and rabbits, frequently under conditions of restraint. Nonetheless they demonstrate quite complicated hippocampal reactions, and a number of cases are indicated in Appendix 6 for which we could not extrapolate directly from the apparent stimulus or response correlate to the functional significance of cell firing in any one area. We also describe there systematic relations between the different components of the hippocampal formation. The picture presented by these simpler paradigms suggests that fairly non-specific information, related to the priming of subcortical response systems, is sent from the medial septum to the hippocampal formation; and that this is more likely in the case of conditioned than unconditioned responses. As would be expected if they are not encoding specific stimulus features of the environment, medial septal neurons recorded in a non-spatial working memory task have activity which is occasionally correlated with stimulus presentation, but more often correlated with response emission or reward delivery; and 'incorrect responses are not associated with activation, indicating that the medial septal area is only active under conditions in which the appropriate response rule is retrieved for a given stimulus' (Givens 1996a; see also Kita *et al.* 1995).

Where reinforcement is present, the simpler paradigms suggest that the non-specific septal input and specific entorhinal input are combined to produce CA3 and CA1

responses. Consistent with a combinatory function of this kind, there are hippocampal cells with highly specific, apparently 'relational', fields. This specificity (within any single task) extends even to apparently simple tasks, such as differential reinforcement of low rates of response (DRL), during which each of many cells has a unique firing field (Young and McNaughton, in preparation). In all these cases, we argue (see Chapter 1) that hippocampal activity reflects relatively simple information about available goals (and hence choices) provided to it by other structures, gated by aminergic inputs which indicate importance, and by mismatch (in either direction) between septal and entorhinal inputs, indicating a need for information gathering. We further suggest (following, for example, Wiener 1996) that distinct fields within a single task reflect distinct components of that task (subgoals, if you will).

With our overview of these data as a foundation, we now turn to observations from more complex tasks; for these, however, there is much less information about the transformations occurring between the different areas of the hippocampus.

7.2 SINGLE-UNIT RESPONSES: CORRELATIONS WITH SPATIAL POSITION

In the studies discussed so far, either the animal was unable to move itself through space or, in the more complex experiments, the tasks (or at least predictors of reinforcement) have been non-spatial and, in our summary, we have ignored the possible relationships between cell firing and spatial position (although in fact, in the latter experiments, much of the task-related firing cannot in any case be accounted for by correlations with spatial position). In a complex spatial environment, however, two main firing patterns are observed, both of which have a connection with place.

Some hippocampal units fire principally when the animal is moving from place to place, without any clear relationship to its mode of movement or destination. Their firing has a high rate. It is strongly correlated with movement theta (see below) and the relevant units often fire in bursts phase locked to theta (Feder and Ranck 1973; Ranck 1973; O'Keefe and Nadel 1978). As movement speed increases, previously non-rhythmic units become rhythmic and previously rhythmic units increase their rhythmicity (Rivas *et al.* 1996), so it is not clear whether units with these characteristics constitute a categorically distinct class of neuron or a set of neurons temporarily showing a particular mode of responding. O'Keefe and Nadel (1978) term the movement-related cells 'displace' units; they are identical to the 'theta units' in other reports (O'Keefe 1976).

By contrast, other hippocampal units appear to fire only when the animal is in a particular place. Descriptively, the unit has a 'place field'. As a result, they are often called 'place cells' (see O'Mara 1995; Wiener 1996 for reviews). Although these cells do not show individual theta activity, the individual probability of their firing is related to the ongoing theta rhythm, and their average firing as a population shows theta modulation (e.g. Skaggs *et al.* 1996, Fig. 2). The description of such spatially tuned units by O'Keefe and Dostrovsky (1971) provided the main impetus for the spatial theory of hippocampal function, although similar suggestions have been made on the basis of lesion experiments (e.g. Olton and Isaacson 1968; Mahut 1971). The early work on 'spatial

fields' of hippocampal cells has been reviewed by O'Keefe and Nadel (1978). In 1991, the journal *Hippocampus* (pp. 221–92) published a forum 'Is the hippocampal formation preferentially involved in spatial behaviour?', from which it is clear that there is still substantial room for interpretation of the data on this issue.

The basic phenomenon of place fields is not in question. What we must consider here is its generality, the extent to which the name applied to the field matches the functional significance of the firing and, most importantly, the extent to which a cell which has a place field can be called a place cell. We shall argue that 'place cell' is a misnomer and that 'place field' is descriptively, but not functionally, accurate.

There are many different properties of place fields which we consider below. A common feature, important for our theory, is this: the data on place fields suggest that 'hippocampal neuronal activity may reflect the association of movements with their spatial consequences' (Foster *et al.* 1989, p. 1580), and they support the view that the hippocampal formation is involved in path integration (Wiener and Bertholz 1993, cited by Wiener 1996; B. McNaughton *et al.*, 1996; see Wiener and Korshunov 1995). This notion—that a spatial field is largely the result of the association of a particular class of movement with a particular consequence—is similar to the idea of an 'available goal' field (see Chapter 1) which we used above in relation to the simpler paradigms. Certainly, in everyday speech, a goal is often a place.

The specifically spatial basis of place fields was demonstrated by O'Keefe and Conway (cited in O'Keefe and Black 1978) using the specially constructed environment shown in Fig. 7.1. Animals were trained inside a square enclosure.

> The walls were formed by floorlength black curtains. Within the enclosure, there were four cues by which the rat could locate itself: a dim white light, a white card, a buzzer and a fan. Four male hooded rats were made hungry and trained to go to one of the arms of a T maze to obtain food. From trial to trial, the maze and the cues were randomly rotated by some multiple of 90° relative to the environment, but maintaining the same spatial relationship to each other. In order to rule out other intra-maze cues as a means of solution, the arms were interchanged from trial to trial. Body turns were ruled out by randomly rotating the stem of the T maze 180° relative to the cross bar so that on one half of the trials a right turn was required to reach the goal while on the other half, a left turn was required. After the rats had learned the task they were further trained so that after running to the goal arm and consuming the reward they should run to the non-goal arm and thence back to the start arm, where they received a second reward. Thus on each trial they made a complete circuit of the maze, giving the hippocampal units an equal opportunity to fire on all parts of the maze. The place where the units fired on the maze was recorded by pulsing a light-emitting diode on the rat's head whenever an action potential occurred and photographing these light flashes with an overhead camera. (O'Keefe and Black 1978, p. 184.)

Using this paradigm, O'Keefe and Conway were able to determine that a number of cells fired in response to particular relations between the four controlled cues inside the enclosure (light, card, buzzer, and fan). An example of such a place field is shown in Fig. 7.2. On probe trials, one or more of these cues was removed. Some place units responded, as in the example shown in Fig. 7.2, by extending their firing field into the entire test environment; others responded by ceasing to fire. In either event, the place field was lost. The increased fields of some cells may be related to the behaviour of the

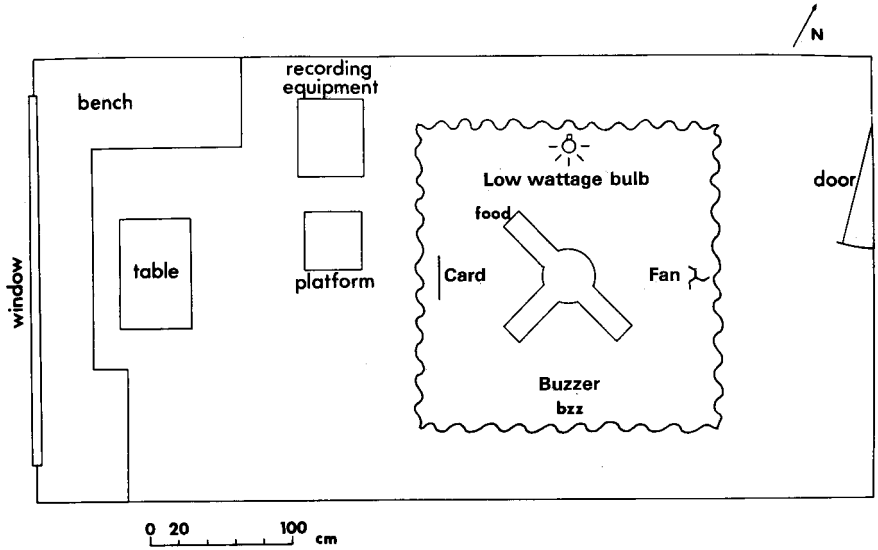


Fig. 7.1 Layout of the experimental room in O'Keefe and Conway's single-unit experiment. The cue-controlled enclosure is the curtained-off area just to the right of centre. (From O'Keefe and Black 1978.)

'misplace' units also described by O'Keefe and Nadel (1978), that is, units which fire when the rat goes to a place and fails to find a stimulus (e.g. light or food) which has previously been there (O'Keefe 1976). Both place and misplace units have been described by other workers (Ranck 1973; O'Keefe and Nadel 1978, p. 209; Olton *et al.* 1978a).

Let us now look more closely at the properties of the place fields. Note, first, that there is something more than a little miraculous about constructing an ad hoc environment, sticking a wire into the middle of the brain and immediately encountering a place field. Obviously the hippocampus cannot come prewired to respond specifically and uniquely to, say, 'a fan equidistant from a buzzer and a light, and 90° round from the light'; and yet for some units the place field appears, as far as can be told, *immediately* the animal encounters a new apparatus (Hill 1978). It could be that there are many cells closely packed together and those not related to the current place are silent (and, of course, the experimenters keep searching until they find cells which are firing at some time in their apparatus). However, the same cell can be shown to have a place field in each of a number of different environments. Interestingly these multiple fields 'were very different in the different environments. . . . A single neurone could have a large place field in one situation and a small place field in another situation. In the same situation a neurone can have single or multiple place fields. This could vary between situations. There was no constancy of the place fields of the same neurone with respect to distal cues [common to the environments]. We have not been able to discover any organizing or over-riding principles for the positioning of a single cell's place fields across environments' (Kubie and Ranck 1983, p. 438; see also Wiener 1996). In monkeys, Cahusac *et al.* (1989) noted that different cells had spatial correlates in two different spatial tasks;

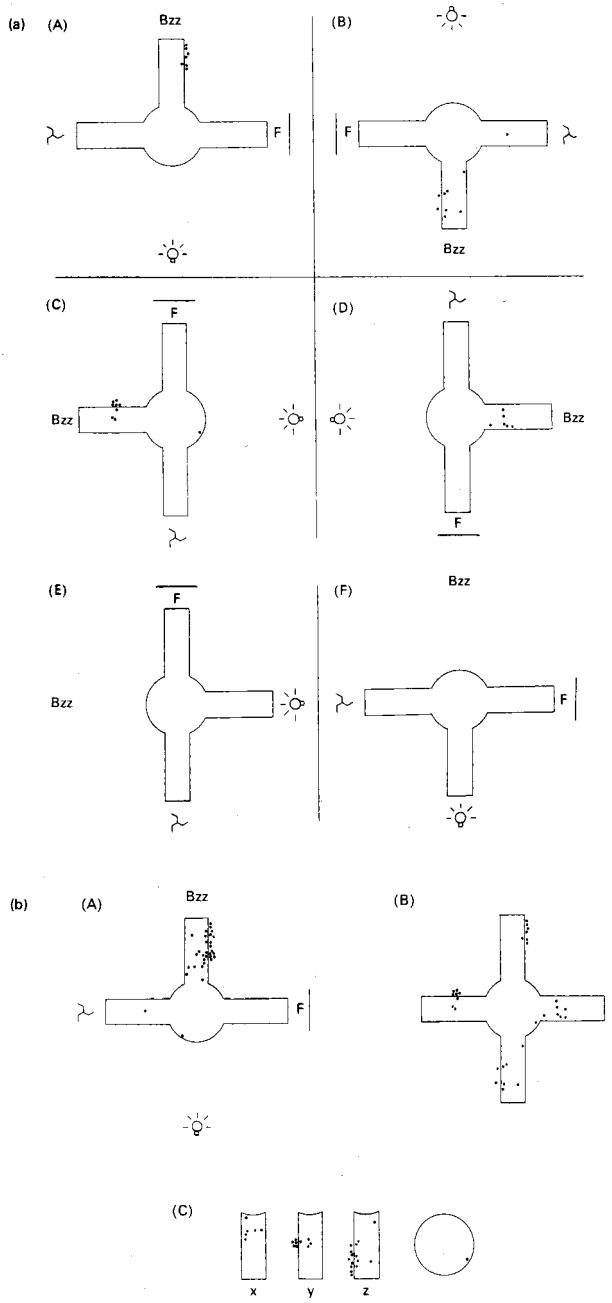


Fig. 7.2 (a) The firing of a place unit when a rat is on the T-maze inside the cue-controlled enclosure (see Fig. 7.1). Each dot represents one action potential. Four trials are shown in A–D in which the T-maze and the cues on the wall have four different orientations relative to the external world. The orientation of the external world is constant. Note that the unit fires when the rat is in the start arm when it is on the side close to the buzzer, regardless of the orientation relative

and there have long been data to show (e.g. Ranck in Elliott and Whelan 1978, p. 310) that, in rats, the same cell can have one correlate in a clearly spatial task (the radial arm maze), another in a task which may or may not be spatial (retrieving pups in a nest box), and yet a third in a task which is almost certainly not spatial (bar pressing on a DRL schedule).

Second, Vinogradova's (1975) experiments on habituation to simple sensory stimuli and Berger and Thompson's (1978a,b) on conditioning of the nictitating membrane response, mentioned above (see also Appendix 6), demonstrate 'fields', but under conditions which are likely to inspire an interest in space in only a very contemplative rabbit; while rats, in tasks which do allow for spatial coding, nonetheless have many cells with 'cue sampling' and 'goal approach' fields which are orthogonal to spatial location (Eichenbaum *et al.* 1987). In a lever-pressing procedure, which of course has spatial features, if the schedule is DRL so that the problem requires an essentially non-spatial solution, virtually all cells have non-spatial fields (Young and McNaughton, in preparation). So, while the firing of a hippocampal unit can indicate a rat's position within a single relatively unchanging environment (Wilson and McNaughton 1993), it cannot tell us unambiguously where the animal is under all conditions. Thus, hippocampal cells, even those that can under the right conditions earn themselves the title of 'place cells', are flexibly and only temporarily allocated to the representation of a range of environmental features, of which spatial location can for a while be one.

Third, even under experimental conditions in which location appears to be a necessary condition for firing, it is not a sufficient condition for firing in the place field. A unit will often fire only in that subset of cases when a particular organization of the external stimuli is viewed (e.g. McNaughton *et al.* 1983; Breese *et al.* 1989; see also Rolls and O'Mara 1995). Furthermore, if a source of reinforcement is placed within a place field, then the place cell can be silenced (Gothard *et al.* 1996). Why this should be we shall consider shortly in relation to the apparent task-dependence of spatial fields.

Fourth, no-one has ever reported anything approaching a topographic map of spatial fields within the hippocampus. This failure contrasts, for example, with the allocentric and body part-related maps found in areas such as the basal ganglia and frontal cortex, and the visual field-related maps found in visual cortex. There is, in addition, enormous variation in the size and shape of spatial fields from cell to cell in the hippocampus (Muller *et al.* 1987), as well as from time to time in the same cell. Despite this variation, pairs of cells with related firing patterns show 'the same interrelations with respect to place field overlap and cross-correlations, irrespective of actual field location' (Hampson *et al.* 1996), in contrast to groups of *neighbouring* cells, which may show considerable discordance as to whether they are tuned to local or distal cues (Tanila *et al.* 1997).

---

to the external world. E and F show two trials with the start arm rotated 180° so that it is on the side close to the light. There is no unit firing in the start arm. (b) The same unit as in (a), showing the results of trials A–D superimposed in different ways. (A) Cell firing aligned in relation to the controlled cues on the walls; (B) cell firing aligned in relation to the external world; (C) cell firing aligned with respect to the individual specific components making up the maze. Note that the z-arm was used as the start arm twice while the other arms were used only once each. (From O'Keefe and Nadel 1978.)

Fifth, spatial fields often seem to cluster on important locations and to change if the importance of the location changes. There is, to begin with, some alteration and sharpening up of fields during the first few minutes of experience with an apparatus (Muller *et al.* 1987). In one experiment (Breese *et al.* 1989), spatial fields clustered on five locations at which water was randomly delivered and tended to omit locations where water was not delivered. Delivering water to only one of the five locations caused most cells (40/47) to shift their field to that location or to change their fields in other ways. The fields appear to relate, then, to task demands. In a 'foraging task in an open field, . . . hippocampal CA1 pyramidal cells discharge when the rat visits the firing field, regardless of the direction that the animal is facing' (Wiener 1996, p. 341), whereas 'in tasks involving repeated, stereotyped trajectories between fixed locations, hippocampal place fields are highly directionally sensitive' (B. McNaughton *et al.*, 1996; Markus *et al.* 1995). Similarly, O'Keefe and Speakman (1987) tested the effects of rotating a set of spatial cues, which signalled reward location, with respect to the global environment, which did not predict reward location. Most cells (33/55) had spatial fields fixed with reference to the cue array, while a more modest number (15/55) had spatial fields fixed with reference to the global environment. Of particular interest is the fact that, if the cues were removed and a spatial field remained in a previously cue-related unit, this spatial field predicted the animal's choice even if this turned out to be an error. Remarkably, exploration of a range of novel environments can sometimes drastically alter the distribution of place fields in an apparatus in which the rat performs a regular task that is the same before and after exposure to the novel environments (B. McNaughton *et al.*, 1996). As mentioned previously, where cells are identified which fire in relation to a movable food source and other cells have place fields, the placement of the food source in the place field silences the place cell (Gothard *et al.* 1996). In all these ways, then, the firing of these cells is determined by spatial location only in interaction with other environmental features that are of motivational significance for the animal.

This line of argument should not be taken, however, too far: space remains of the greatest importance in determining hippocampal unit responses. Spatial fields are not directly related to reinforcement as such, and are usually sufficiently well spread within a single environment to allow accurate prediction of the animal's position (Wilson and McNaughton 1993). In the monkey, while extensive work has reported many detailed results with only modest evidence of any spatial biases in firing (see O'Mara 1995), this is most probably because the monkeys cannot move and the experimenters carefully excluded spatial position as a possible correlate of the monkey's goals. When Cahusac *et al.* (1989) included spatial position as a factor in determining correct responses, they found a number of neurons with spatial correlates; and, when Rolls and O'Mara (1995) moved a monkey about with a robot, they found cells which fired in relation to the specific place at which the subject looked and in relation to movement to a place, as well as in relation to other more complex conditions.

Sixth, distortions of the size and shape of the arena in which the animal is placed do not produce equivalent distortions in the spatial fields of all cells, with the single exception of rectangular arenas which vary in side length (O'Keefe and Burgess 1996). The firing pattern in a rectangular arena could not be predicted from the firing pattern in a circular arena (Muller and Kubie 1987). Importantly, if a transparent barrier was placed

so that it intersected part of a spatial field, it could destroy that field (or sometimes expand it), whereas the lead base used to hold the barrier did not have this effect when used without the barrier (Muller and Kubie 1987). This suggests that spatial fields result not from the animal's being in the location (which would be easily discerned from cues visible through the transparent barrier), but from the animal's knowledge of what it can do at that location (which would be changed by the barrier). Consistent with this inference, there appears to be a memorial component to spatial fields, so that the cell will maintain its place field even in the dark, provided the animal knows its starting location (see B. L. McNaughton *et al.* 1989; Markus *et al.* 1994). However, direct placement into a dark environment (eliminating starting location information) can cause a *remapping* of the cell's previous place field which is then *maintained* when the light is put back on (Quirk *et al.* 1990). Similarly, conflicting vestibular and visual information can cause 'unpredictable changes in firing characteristics, so that cells either stopped firing, or developed place fields that were altered in overall size, shape, and eccentricity' (Sharp *et al.* 1995, p. 173). Thus, we again see that allocation of a particular cell to a particular representation (or what the experimenter takes to be such a representation) depends upon a combination of factors, not spatial location alone, and can change rapidly.

It follows from all of the above that 'place field' is a convenient descriptive term within specific experimental paradigms rather than being functionally accurate, and that the term 'place cell', in any literal sense, is a misnomer. If a cell has a place field in any particular environment at any particular time, this does not necessarily and uniquely indicate that the animal will be at that same point in allocentric space when the cell fires at any other time. There appears to be no appropriate spatial mapping in the hippocampus to allow decoding. The position (and indeed spatial nature) of the field can change dramatically in a spatially unchanged environment when response tendencies change. With cues which provide an additional spatial-like frame of reference, some cells code geographical space, some the cue space, and the majority code whichever frame of reference predicts reinforcement. In two or more shapes of apparatus, there is no obvious allocentric relation between the positions of spatial fields in one compared to another.

If place fields do not solely reflect spatial position, how can we account for them? To answer this, let us first look at the theoretically simpler head direction fields which have been found in areas connected to the hippocampal formation.

7.3 THE RELATION BETWEEN SPATIAL FIELDS AND HEAD DIRECTION FIELDS

There is evidence which suggests that the spatial fields of hippocampal units could be constructed by 'head direction' information, which is in turn derived from visual (landmark) input as well as vestibular and proprioceptive (idiothetic) input (for a review of head direction cells see Taube *et al.* 1996).

Visual input to the head direction system is most likely to come from 'intermediate/deep layers (IV–VII) of superior colliculus [which] play a special multisensory role in spatially guided behaviour . . . [and] project to the lateral dorsal nucleus of the thalamus

(LDN), which also receives direct input from retinal recipient pretectal nuclei. . . . LDN cells . . . in turn project directly to a number of cortical structures, notably the hippocampal formation' (Mizumori and Williams 1993). These LDN cells 'selectively discharged when an animal's head was aligned along particular directions in space, irrespective of its location in the test room. . . . If the light was . . . turned off, directional firing was maintained briefly . . . [and then] appeared to systematically 'rotate' in either the clockwise or the counterclockwise direction. . . . Maximal directional firing was achieved only when the rat viewed the entire test room, and not just the scene associated with the directional preference of the cell' (Mizumori and Williams 1993). Up to a minute of exposure to the environment could be needed for the directionality to appear. LDN, therefore, appears to acquire and then retain (like place fields, even in the absence of visual input) preliminary positional information which is head centred. Critically, 'inactivation of the LDN not only impairs spatial performance in rats, but it also disrupts location-specific discharge by hippocampal cells' (Mizumori and Williams 1993; Mizumori *et al.* 1994).

Mizumori and Williams (1993) and Taube *et al.* (1996) suggest that the information from the LDN is sent to the post-subiculum (as well as directly to the entorhinal cortex and to the subiculum). The post-subiculum receives a large number of other inputs, but appears to have relatively similar 'head direction' cells to those in LDN (Taube *et al.* 1990a,b). However, the directionality of these cells is not affected by LDN lesions (see Taube *et al.* 1996). This is surprising, since hippocampal spatial fields are affected by such lesions, and suggests that LDN information taking the direct route to entorhinal cortex is crucial for the hippocampus but only incidental for the post-subiculum, which depends on an intact anterior dorsal nucleus of the thalamus for its crucial directional information (Fig. 7.3). The anterior thalamic cells appear to predict future head direction rather than to indicate current head direction (Blair and Sharp 1995), and so it is possible that the post-subiculum receives both current and future head direction information.

This stimulus-oriented description is easy to follow, but we should note that the source of the original information, the superior colliculus, is much more concerned with spatially guided behaviour (and, indeed, affective significance; Redgrave and Dean 1991) than visual space itself. For example, in discussing motor control in frontal cortex, we made the suggestion that the frontal eye fields are an area of pre-motor processing which, of necessity, engages the eyes via the superior colliculus in order to provide the information required for the upcoming movement. Indeed, Mizumori and Williams (1993) suggest that 'the LDN may be part of a neural system that subserves directed attention to constellations of salient cues in the animal's visual world. In this context it is worth noting that the structurally similar thalamic nucleus in primates, the pulvinar, has been shown to play a significant role in visual spatial attention processes. Specifically, Desimone and colleagues suggest that the pulvinar may *gate extrastriate responses to distracting visual input, thereby focusing one's attention onto a visual target*' (our emphasis). Thus LDN may be concerned with the inhibition of eye movements to inappropriate objects of visual attention ('visual goals').

Taube *et al.* (1996) conclude that a 'motor efference copy' is important in generating head direction cell discharge, but that sustained discharge when the animal is stationary is not consistent with this idea. However, Knierim and McNaughton (commentary

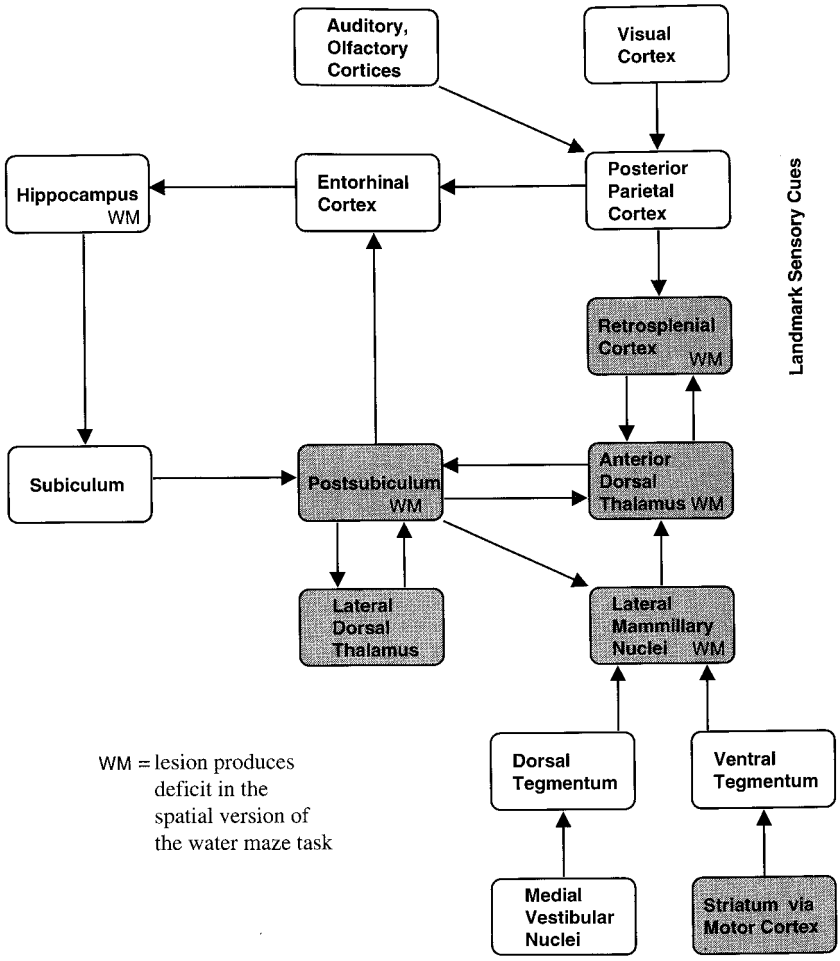


Fig. 7.3 Locations in the brain from which head direction cells have been recorded (shaded boxes) and locations at which lesions have been shown to impair spatial learning (WM). The figure also shows the connections between these areas and the hippocampal formation and areas which are likely to provide the sensory information driving the head direction cells. (Adapted from Taube *et al.* 1996, with the addition of shading of the lateral mammillary nuclei; Taube, personal communication.)

following Taube *et al.* 1996, p. 484; see also Knierim *et al.* 1995) note that 'both place cells and head direction cells do maintain their tuning properties when the animal is passively rotated, as long as the animal is not tightly restrained. It thus appears that some type of 'motor set'—the preparedness for motion—has an influence over the responses of these cells, rather than any copy of actual motor commands'; and, as we quoted earlier, that 'hippocampal neuronal activity may reflect the [learned] association of movements with their spatial consequences' (Foster *et al.* 1989, p. 1580; see also O'Mara *et al.* 1994 on the effects of passive movement of a monkey on hippocampal cells).

7.4 SPACE, DISCRIMINATION, CLASSICAL CONDITIONING, AND HABITUATION

Let us now draw some parallels between the properties of spatial and non-spatial fields. First, a stimulus-oriented description has often been the easiest way to describe the firing correlates of a unit; however, the response elicited by the stimulus has also to be taken into account. Two examples illustrating this point are that: novel stimuli that are effective in triggering hippocampal cells to fire appear to be characterized not so much by novelty *per se*, but rather by elicitation of an orienting reflex (Appendix 6, Sections A6.2.1.8, A6.2.2); and unit activity modelled the nictitating membrane *response* in the case of classical conditioning to such an extent that, initially, it was thought that the hippocampus was the 'neuronal substrate of classical conditioning' (Berger and Alger 1976).

Second, responding *per se* is not a critical determinant of the fields: there may be no firing in relation to an unconditioned response; cells fire during a delay even when the upcoming response is indeterminate; similarly, while place fields disappear when responding is impossible, they, nonetheless, do not require the actual making of a response for their presence.

Third, fields are sensitive to the importance of stimuli: in areas CA3 and CA1 firing to simple stimuli swiftly habituates, whereas firing to stimuli which predict an important event like reward is incremented (Section 10.3); the position of a 'place field' is occasionally shifted to a completely different part of the apparatus if the delivery of reward becomes concentrated there; and field directionality and spatiality depend on task demands.

Fourth, even a clear response correlate of CA1 or CA3 firing (e.g. the modelling of the nictitating membrane response to the CS; Berger and Thompson 1978a,b) need not predict functional involvement in the behaviour as such (conditioning of the nictitating membrane response has been shown to occur outside the hippocampal formation, and removal of the hippocampus did not affect this conditioning). The same appears true for spatial fields. These are very clear in runways, yet learning to run down the runway is unaffected by hippocampal lesions.

Fifth, firing in areas CA1 and CA3 appears to result from integration (of different sorts under different reward conditions) of information arriving from both the septum and the entorhinal cortex. That the entorhinal input to the hippocampus is altered by additional information is shown by the fact that CA1 and CA3 units have much sharper spatial fields than do entorhinal units, and show different field changes when the environment is modified (Quirk *et al.* 1992). That the non-entorhinal input must contain information equivalent to that of the entorhinal input with which it is integrated is shown by the fact that, after entorhinal lesions, hippocampal place-like fields remain. However, the information of non-entorhinal origin is only loosely similar to entorhinal information, since these place-like fields are linked to local rather than distal cues. The additional information is most likely to come from the septum. Fornix lesions have effects similar to those of entorhinal lesions in that they do not eliminate all spatial fields (Miller and Best 1980). The normal spatial fields of CA1 and CA3 appear to depend, then, on the combination of somewhat different, but congruent, types of information from the septum and the entorhinal cortex which, when they interact, produce tighter

spatial fields than can either input alone. That this interaction occurs within the hippocampus proper is suggested by the fact (Kita *et al.* 1995) that, in monkeys, lateral septal neurons (which receive output from area CA3) are more likely to have spatially related fields than are medial septal neurons (which provide input to area CA3). The effects on hippocampal firing repertoires of the septal or entorhinal lesions may, then, be the simple result of a reduction in the amount of information available to areas CA1 and CA3 for their computations. The lesion-induced effects are similar to the changes produced by blinding and deafening the rat (Hill and Best 1981), and also appear to match the effects on spatial fields of testing rats in the dark (see Markus *et al.* 1994). These data all suggest, then, that CA1 and CA3 place fields depend on septo-entorhinal interactions in much the same way as we have described for non-spatial fields (Section 10.3), and that loss of either input coarsens a property akin to sensory acuity.

This consistency and the principle of parsimony give us some justification for our presentation of hippocampal fields as signalling exactly the same general class of feature under all circumstances, spatially complex or not; with the variations in observed correlates determined by the structuring of goals (or subgoals) within each paradigm. We agree with Wiener (1996, p. 355) that 'the discharge correlates for location and for task-related behaviours appear to be extremes of a continuum of responses in a single population of neurones. In intermediate cases, the discharge correlate is for conjunctions or recombinations across certain information domains, while generalizing over others. These responses seem to originate in a single population of neurones since cells with location-selective discharges can, in different circumstances, also have task-related behavioural correlates.'

7.5 PATH INTEGRATION

'In spite of O'Keefe's original suggestion that the hippocampal navigation system might be fundamentally organized in terms of the integration of self-motion, most studies [have assumed] . . . that the primary source of spatial information . . . is the input from visual landmarks and other stimuli in the environment. . . . A growing body of evidence suggests that the hippocampal formation is a fundamental component of [a] path integration system . . . based on a self-motion metric' (B. McNaughton *et al.*, in press; see also Wiener and Berthoz 1993, cited by Wiener 1996; B. L. McNaughton *et al.* 1991, 1997).

Path integration, that is the maintenance in memory of a fixed location through the process of dead-reckoning, in turn, could be a special case of the sequencing of successive goals to achieve some final objective (Wiener and Korshunov 1995), since:

> hippocampal neurones discharge selectively as the subject . . . responds during successive phases of a behavioral task [to] stimulus and reward presentations as well as task-related behaviors . . . [so that] *all of the behavioral components of the experimental task can be comprehensively represented.* [For example, in] a water search task . . . [specific] neurones discharged when the rat performed [specific] task-related behaviors at several locations in the arena, each of which were virtually identical both in appearance and in significance with respect to the task requirements. . . . Thus the motor components of these behaviours were not the principal correlate of the unit activity. The fact that some of these neurones were selective for either trajectories towards the corners or

the center also argues against a simple relation to sensory cues or rewards. . . . It appears then that this discharge selectivity depended on the behavioral context, that is, the rat being engaged in only one of the phases of a cycle of actions that it had adopted to satisfy the exigencies of the task. (Wiener 1996, pp. 346–8, our emphasis.)

This seems true even with nictitating membrane response conditioning, since a recent study of trace conditioning (a hippocampal lesion-dependent task) also found specific hippocampal cells firing in relation to specific components of the task, with essentially all components of the task covered (Weiss *et al.* 1996); and, as we note in Appendix 6, the same appears true of at least some pyramidal cells even in the simple conditioning of this response (Berger *et al.* 1983).

Let us look closely at the idea of a hippocampal cell being activated by a specific goal within a sequence of task-related goals. For both the non-spatial and spatial cases, we found that two units with apparently the same type of correlate under one set of conditions could have different correlates from each other when the conditions changed. Also, spatial and head direction fields appear to be seen, not when the animal is about to move, but whenever it could do so. The firing of an individual cell must, therefore, be related to the available goals, with many such being signalled in the hippocampus simultaneously. For example, cells which signal the presence of a left-hand goal and cells which signal the presence of a right-hand one will fire simultaneously when the subject reaches a T-junction. This fits with the fact that changing the location of reward changes the position of the place fields of only some cells. The goal fields have remained constant and the change in location of reinforcer has shifted the position of goals related to that reinforcer, but not goals related to other reinforcers. The notion that hippocampal cells are activated by available goals is consistent also with the one case where place fields seem predictable from apparatus to apparatus: a set of various-sized rectangular boxes. Here, the available goals (or paths to be integrated) should be strongly related from apparatus to apparatus. The idea of 'available goal' fields can also account for the lack of topographic mapping of spatial fields and the fact that, when the location of spatial fields is altered, closely related fields tend to stay together and to show similar changes in size (Hampson *et al.* 1996).

There are no doubt many possible models of path integration and many possible hypotheses as to where in the brain the actual integration is carried out. Such models may or may not involve a true 'cognitive map' in O'Keefe and Nadel's (1978) sense. But, whatever the precise means used by the hippocampus to obtain information about the alternatives which can be chosen at any particular point in space or time, such information should allow it to 'select and test hypotheses, or strategies, concerning the solution [to a problem]' (O'Keefe and Nadel 1978, p. 2). Such selection and testing by the hippocampal system would be particularly important when a novel short-cut between two points can be substituted for a prior longer route by comparison of alternative immediately available goals. It has been argued by Bennett (1996) that the capacity to select short cuts in this way can be computationally simple and can account for all the current evidence for 'cognitive maps' of the type proposed by Tolman and by O'Keefe and Nadel (see also B. L. McNaughton *et al.* 1991). If so, this brings our account of hippocampal fields very close indeed to that required by O'Keefe and Nadel's (1978) cognitive mapping account.

7.6 TOPOGRAPHIC MAPPING OF SINGLE-CELL FIELDS

As we have seen, place fields are not topographically mapped into the hippocampal formation. Yet when the location of spatial fields is altered, closely related fields tend to stay together and to show similar changes in size (Hampson *et al.* 1996). This suggests that there is non-spatial topographic organization in the hippocampus.

We certainly have reason to look for some kind of topography. More and more brain areas are being shown to represent information topographically. The afferents, efferents, and internal structure of the hippocampus are themselves strongly topographically ordered (see Appendix 4). We also have a ready candidate for the type of entity which might provide the basis for topographic mapping. We have suggested above that the receptive fields of hippocampal cells relate to 'available goals'. It follows that the hippocampus should contain a map of 'goal space'. Less obvious, however, is on what psychological dimensions available goals could be simply mapped. Our way of dealing with this problem is to look at a likely mapping in neural terms, leaving open the question whether the result will be amenable to ready psychological description. The mapping we propose, then, derives from the hypothesis (and from the evidence reviewed above which supports it), that hippocampal cells code for available goals. This hypothesis implies that, when such cells fire, there will be activity in one or more of a number of distinct goal-processing areas in the brain that project to the hippocampus, or in separate modules within such an area. It is a brief step, then, to suggest that the inputs to, and outputs from, the hippocampal formation are topographically mapped in terms of these goal-processing areas. One possibility, for example, is that hippocampal cell firing results from its receipt of an efference copy of output which, in the absence of hippocampal functional output, can enable particular motor programs by way of projections to motor programming areas.

This suggestion, that the hippocampus represents available goal fields topographically so that its output can target specific goal-processing systems, is consistent with Risold and Swanson's (1996, p. 1484) conclusion 'that different hippocampal regions map in an orderly way onto hypothalamic systems mediating the expression of different classes of goal-oriented behaviour'. Our hypothesis predicts, indeed, that similar topography will be apparent in the input and output connections of the hippocampal formation with other areas. Such evidence as there is (Appendix 4; e.g. the strong relation of amygdaloid connections to the temporal hippocampus and to the subicular–CA1 border), is consistent with this prediction. Notably, many of the major targets of hippocampal efferents are areas concerned with planning (prefrontal cortex, cingulate cortex), response organization (amygdala, accumbens, basal ganglia, hypothalamus), or internal adjustments related to the anticipation of action (amygdala, hypothalamus).

There are also indications that the topographic mapping of goals *within* the hippocampal formation may ultimately be understandable in terms of a dimensional organization. We have already noted that some goals are more easily distinguished in response terms (approach–avoidance) and others in stimulus terms (left lever–right lever), although the very nature of a goal implies that each must have an admixture of both stimulus and response qualities (Fig. 1.7). In Appendix 4, we suggest that septal and temporal hippocampus receive more dorsal trend, 'where', information and more ventral

trend, 'what', information, respectively (see also B. L. McNaughton *et al.* 1989). We also note that the amygdala (concerned as it is with the affective aspects of goals, i.e. what they are like rather than where they are) has stronger connections to the temporal hippocampus than to the septal hippocampus. Consistent with these principles of topographical organization, as one moves from the septal to the temporal part of the hippocampus, there is a decrease in the numbers of cells with spatial fields (from about 50 per cent to 20 per cent) and a concurrent increase in the size of such fields as are detected (Jung *et al.* 1994; but see also Poucet *et al.* 1994), suggesting again a more 'where' (dorsal trend) orientation of septal fields and a more 'what' (ventral trend) orientation of temporal fields. This possible dimension (what–where) of the septal–temporal lay-out of the hippocampal formation then maps topographically to dorsomedial–ventrolateral in the septum and caudal–rostral in the hypothalamus (Fig. 9.5, p. 220; Risold and Swanson 1996).

Note that the septal–temporal dimension appears to be truly dimensional and not, as would be implied by the dorsal–ventral distinction, categorical. The 'where–what' separation within the hippocampus reflects variations only in the *proportions* with which dorsal and ventral trend information arriving in the entorhinal cortex is mixed as the two streams are transferred to the hippocampus. It is in the nature of the concept of a goal that it must combine 'what' and 'where' to some extent; hence, perhaps, the otherwise peculiar anatomical fact that the hippocampus (which can be thought of as a remnant of the primordial evolutionary and developmental origin of the dorsal trend) receives a confluence of both dorsal and ventral trend information.

What of the orthogonal dimension (CA3–CA1–subiculum) in the hippocampal formation?[1] In discussing the habituation and simple conditioning results of single-unit recording experiments above (and see Chapter 10 for greater detail), we suggested that there is a fairly complex set of logical gates which takes a relatively non-specific septal signal and combines it with a specific entorhinal–dentate signal and with aminergic input to determine the extent to which each stage of the trisynaptic pathway will be activated, and the extent to which such activation has functional consequences. Place fields, on this scenario, are recorded with both the dentate–CA3 and CA3–CA1 gates open. The effects of lesions on place fields suggest that septal input relates to the local presence of important stimuli (and hence rather gross response tendencies, consistent with the multimodality found in septal firing repertoires in the simpler single-unit recording paradigms), while entorhinal input, once processed in the dentate, provides more specific positional information (and hence relates to the more detailed organization of response programs, consistent with the modality-specific firing repertoires of entorhinal cells found in the simpler paradigms). We presume that CA3 activity is the result of the integration of dentate with septal input. Consistent with this assumption, as we go from dentate to CA3 (Jung and McNaughton 1993) to CA1 and then subiculum (Barnes *et al.* 1990; Sharp and Green 1994), spatial specificity decreases. This second dimension of hippocampal mapping may then reflect 'level of processing' in some sense or other.

1. While we talk about CA3, CA1, and subiculum as three distinct fields, the mapping may be more continuous than this: area CA2 has long been recognized as potentially distinct; CA4 may be an extension of CA3 (which itself has occasionally been divided into subregions a, b, and c); the CA1–subiculum border appears to have a specific relation to the amygdala; and other subregional strips are appearing in the literature.

7.7 CONCLUSION

It is, then, undoubtedly possible to demonstrate correlates of hippocampal cellular firing repertoires that are suitably and succinctly described as place fields. However, the correlates of hippocampal firing are not uniquely spatial. A hypothesis that is able to encompass the bulk of the data from both spatial and non-spatial tasks is that the firing repertoires of hippocampal neurons represent available goals. This hypothesis is one that we shall adopt in the remainder of the argument pursued in this book.

8 Memory and the septo-hippocampal system

In this chapter, we review the role of the septo-hippocampal system in memory and of damage to this system in amnesia. For many people the function of the hippocampal formation is synonymous with at least some type of memory control and so our devotion of only a single chapter to this topic requires an explanation. This is easy to provide. If the behavioural effects of septo-hippocampal system damage are reviewed as a whole (as in Appendix 8), we discover an enormous variety of effects, a number of which are clearly nothing to do with changes in specialized forms of memory. The idea that the hippocampus is primarily a memory control structure can only be maintained if the bulk of this literature is ignored. The facts that the data are not recent and do not fit with 'memorial' preconceptions about the hippocampus are neither of them reasons for implicitly or explicitly (Eichenbaum *et al.* 1994) ignoring them. Indeed, given that the bulk of the non-memorial effects of hippocampal lesions imply changes in more fundamental processes than memory, the onus must be on anyone analysing 'types of memory' to exclude these more fundamental processes by appropriate controls before claiming that this or that favoured test is specific for this or that type of memory. As noted in Chapter 1, the distinguishing feature of our own approach, and the one that might appear most idiosyncratic, is that we do not assign to the hippocampus a specifically memorial function. Rather, we believe that the effects of hippocampal lesions in tests of memory are the result of changes in more fundamental processes that are *not* specific to memory.

However, to most workers, the fact that hippocampal lesions produce 'amnesia', and the fact that long-term potentiation (probably the best current molecular model for memory) was discovered in the hippocampus, make a compelling case for the view that the hippocampus is one of the key memory structures in the brain. To the committed memory theorist, the fact that hippocampal lesions produce many non-memorial effects (namely, the majority of the data reviewed in Appendix 8) will be shrugged off as irrelevant to its memorial function. In this chapter, therefore, we derive our perspective solely from the data on tests of memory. We argue that 'amnesia' does not in fact involve a loss of memory itself, but is better viewed as 'catastrophic hypermnesia' (a failure to suppress the remembering of inappropriate information); that the hippocampus is not a site of long-term memory storage; and that, even if it were, the properties of hippocampal long-term potentiation are not appropriate for such storage.

If we attempted to start with all of the available data and then integrate them, we would need many volumes to complete the task. Indeed, entire volumes have already been written on restricted aspects of memory or of the biology of some such aspect (e.g. Baddeley 1986; Cohen and Eichenbaum 1993; Fuster 1995). Instead, we adopt an approach driven by the theory we present in the next two chapters. First, we provide a critique of a number of recurring myths about the phenomenon of amnesia that have led to considerable confusion in the literature. Second, we offer a summary account of the role of the hippocampal formation in memory and relate this to some of the critical

experimental data in the literature and to the more influential current theories. We shall, therefore, use the existing theories as a means of summarizing the vast body of literature, and our comparative critique of them to highlight the specific experimental data which distinguish our approach from theirs. In this way we hope to achieve an effect equivalent to a comprehensive review of the literature and a mapping of that literature to our theory, but in minimal space.

8.1 HUMAN VERSUS ANIMAL AMNESIA

The first myth we need to dispel (most common in the human cognitive literature) is that the gulf between ourselves and other animals is so great as to make comparison irrelevant. This is a vital myth to eliminate, since the bulk of the data on brain processes underlying memory come from non-human subjects. It has also allowed many theorists to retain concepts which, while adequate in reference to their own data domain and preferred species, are quite inconsistent with data obtained by other methods and from other species.

The idea that human and animal memories are fundamentally different has not, in the past, seemed totally unreasonable. Indeed, at the time of the first edition we remarked that 'there appears to be a considerable difference between the effects of hippocampal lesions in animals and man. In man the most dramatic change is apparently the loss of the ability to form new memories. But we saw virtually no sign of such a change in the animal experiments [we have] reviewed' (Gray 1982, p. 233). Even then, however, the difference was more apparent than real, because, as noted by Tulving (1985, p. 67), animal and human memory research 'seem to represent separate cultures: they begin with different pretheoretical assumptions, they are concerned with apparently different problems, they employ different methods, and they speak different languages. In the animal literature, there are few references to work in human memory; in human memory literature there are even fewer mentions of work in animal learning.' Tulving argued, therefore, for a need to concentrate on the classification of tasks so as to be able to translate between the literatures (see also Roberts 1996; Steckler and Muir 1996). However, this problem need not be solved completely before assessing animal–human differences, especially if we include brain lesions as a means of cross-referencing, since the methods of animal physiological psychology and human neuropsychology differ much less than do human cognitive psychology and animal learning (but see MacPhail 1996 for a non-physiological argument for homology of human and non-human learning).

It is now clear that, provided the methods used to test a non-human subject share conceptually essential features with those used to test a human being, the results are essentially the same. For example, in tests of spatial working memory, rats and undergraduates use similar coding strategies (Kesner and DeSpain 1988) and rats and human subjects with hippocampal damage have similar deficits (R. G. Morris *et al.* 1996). Similarly, rats can show primacy and recency effects (Kesner *et al.* 1994a; but see also Deacon and Rawlins 1995 and references cited therein). Monkeys and human beings show similar use of specific attributes and dimensions in visual discrimination learning and its reversals (Roberts *et al.* 1988); and, provided that they have similar prior experience, lesioned monkeys show 'impairments of object memory . . . analogous to the impairments

of episodic memory seen in human amnesic patients' (Gaffan 1994, p. 305; see also Vnek and Rothblat 1996). Experiments in which human subjects have been tested non-verbally also show that 'animal memory is similar to, at least, the nonverbal part of human memory' (Wright 1994). Likewise, the neural bases of various memory impairments are similar across species (see Kesner 1996). For example, hippocampal lesioned rats and monkeys can learn simple motor tasks like running down a runway or pressing a lever; and H.M., the famous case of bilateral removal of hippocampus, amygdala, and other parts of the temporal lobe, can learn and retain motor tasks such as mirror drawing as easily as unlesioned subjects. Korsakoff patients tested with a standard associative 'animal' task, concurrent object discrimination, show the same severe impairment as do animals (Kessler *et al.* 1986). Prefrontal lesions produce similar dissociations of item and order memory in rats and human beings (Kesner and Holbrook 1987). And, in a final example, when a rat test of non-spatial declarative memory was constructed, it was as sensitive to hippocampal damage as the human tests (Bunsey and Eichenbaum 1996).

However, when amnesics such as H.M. are able to learn and retain certain tasks, they nonetheless cannot remember the task in the sense that they appear unaware, as indexed by verbal report, that they have performed it before or that they are progressively improving. This highlights one of the key sources of confusion in the literature: the testing of human beings via the verbal channel (which is of course not available in animals), coupled with inappropriate inference from verbal results to memory in general. It now seems reasonable to suggest that many such tests of verbal memory are members of a specific subclass of memory tasks which is sensitive to hippocampal lesions in both man and animals, whereas the simplest associative tasks do not belong to this class. (That the verbal channel *per se* is not crucial to the human amnesic deficit, however, is shown by data such as those reported by Kessler *et al.* 1986.)

One must also always be on the look-out for species-specific reactions or bias which can alter the data. For example, verbal rehearsal can provide a human subject with a non-memory strategy to solve a particular problem and hence give rise to quite different results from those seen in the rat, even though the truly memorial performance of both species might be the same. When verbal mediation is excluded, human beings solve radial-arm maze problems in a similar fashion, and with the same implied memory capacity, as do rats (Glassman *et al.* 1994). The need to ensure that memory is the basis of performance or deficit in a task is not confined to experiments with human beings, since, as we shall see, there are a number of rehearsal-like strategies (e.g. maintaining a fixed position within the apparatus in a delayed matching-to-position task) which rats can adopt (but which experimenters usually carefully exclude); these can affect their apparent memory capacity and duration in precisely the same way as does human verbal rehearsal (see also Mumby 1995 and papers cited therein for odour as a possible confound). Related problems arise with the testing of human infants, who can show excellent memorial capacities at a very early age, provided appropriate response requirements are met to permit demonstration of the memories (e.g. Diamond, 1995; Hayne *et al.* 1997; Hayne 1998). In all tests, therefore, one must be careful to exclude species-specific factors before comparing memory systems as such.

Thus, as animal tasks have been adapted more to paradigms typical of human experiments, the animal deficit has come to mimic classic human amnesia (see Squire and

Zola-Morgan 1984). In addition, as the previous 'non-memory' animal data have been subjected to scrutiny, they have often been provided with a mnemonic interpretation (Olton *et al.* 1986). Finally, as human lesions have been found which approximate the discrete lesions made in animals, similar results and similar dissociations have been obtained (Bechara *et al.* 1995).

8.2 THE HIPPOCAMPUS AS THE CRUCIAL SITE FOR AMNESIC DEFICITS

The second myth we must dispel is the idea that the hippocampal formation, and the hippocampus proper in particular, is the primary locus at which damage gives rise to amnesia.

Thus, one source of apparent species difference is the supposed site of damage in amnesic patients, and hence the appropriate lesion for an animal model of amnesia. Here, the nature of the clinical data presents major problems, not least to know what they are. There are three major collections to choose from, each with its attendant sources of confusion. The first consists of observations of a few individuals (such as H.M.) who have suffered extensive controlled damage to the hippocampus as a result of surgery. Reports concerning these individuals stem largely from the late 1950s and 1960s (Milner 1970). There are few new subjects to study, precisely because the huge amnesic deficit described in the early cases more or less rules out this type of surgery for its original application, the relief of temporal lobe epilepsy.

Although there is some evidence that the critical focus for the amnesia indeed lies in the hippocampus (Milner 1970, 1971; Butters and Cermak 1975), all the patients concerned had extensive additional damage to other temporal lobe structures, including the overlying cortex and, at one time thought particularly important, the amygdala. They had also been epileptic for so long before surgery that distal damage resulting from the seizures is highly probable. However, animal studies suggest that the experience of convulsions before hippocampal formation damage does not add to the behavioural deficits observed (Vanderwolf 1995), and O'Keefe and Nadel (1978, p. 416) point out that there are reports of profound amnesia associated with hippocampal damage in the absence of a prior history of seizures. These cases, taken by themselves, provide, however, only limited evidence that amnesia is due to hippocampal damage alone. The small number of patients who underwent the operation and the lack of new cases also make it difficult to use them to test the various new hypotheses concerning the nature of hippocampal function which are still emerging; although H.M., in particular, has by now provided a wealth of data (albeit on a single case who continues to have occasional seizures and remains on anticonvulsant medication). There are, however, four recent patients with lesions (assessed both *in vivo* by magnetic resonance imaging and *post mortem*) mainly restricted to area CA1 who have memory deficits apparently equivalent to those of H.M., but in whom the severity of the amnesia is increased if there is additional hippocampal formation damage (Zola-Morgan *et al.* 1986; Rempel-Clower *et al.* 1996). The data from these patients argue for some truly hippocampal involvement in amnesia.

The second major source of data comes from patients suffering from Korsakoff's syndrome. Since these patients often derive their symptoms from a history of prolonged

drinking, they are in moderately ample supply and likely to continue so. They have, as a result, provided the test bed for many approaches to the amnesic syndrome. However, despite their superficial similarity to the surgical patient group (see Weiskrantz 1978), they do not normally have visible pathology in the hippocampal formation, nor necessarily in structures related to the hippocampus (compare Victor *et al.* 1971; Mair *et al.* 1978). The results of detailed analysis of these patients cannot, therefore, readily be generalized to the surgical group, far less taken as an indication of the effects of pure hippocampal lesions. It is now clear, moreover, that similar, if not identical, amnesias to that seen after hippocampal formation damage are produced by mammillary body lesions (Aggleton *et al.* 1990; Squire *et al.* 1990) and by restricted thalamic lesions (Markowitsch *et al.* 1993; Daum and Ackermann 1994; Parkin *et al.* 1994). The thalamic lesions probably require damage to more than simply the input to the prefrontal cortex from the mediodorsal nucleus (see Markowitsch 1982), and so the amnesia may result from damage to multiple systems. Korsakoff patients can, therefore, be ruled out as representative human *hippocampal* material, whatever their interest with respect to amnesia in general, and whatever links one might subsequently make between hippocampal and mammillary body damage on anatomical or theoretical grounds.

The third major source of data is patients suffering from senile dementia of the Alzheimer type (SDAT). There are some indications that, given sufficient age, the bulk of the population will contract SDAT. The supply of subjects is therefore large and increasing, and, given the obvious importance of SDAT for health policy, much effort has recently gone into analysing its biochemical and genetic causes and assessing therapies. From our point of view, SDAT is slightly more promising than Korsakoff's syndrome in that the hippocampus and other temporal lobe structures are a prime site of degeneration. However, SDAT involves diffuse, and sometimes extensive, damage all over the brain. So, specific conclusions about hippocampal involvement are impossible, especially since animal data have shown that cortical structures adjacent to the hippocampus make a major contribution to amnesia. It is also likely that a primary source of the memory deficit in SDAT, as opposed to patients with damage to the temporal lobes, is loss of cholinergic function. SDAT certainly cannot, therefore, be viewed as an example of pure hippocampal damage. Because of the unsatisfactory nature of all of these sources of data, we shall not attempt a detailed review of the clinical literature here (for reviews at the time of the first edition see Butters and Cermak 1975; Weiskrantz and Warrington 1975; Weiskrantz 1978; Baddeley 1981; and for a more recent review see Squire and Zola-Morgan 1991). In the brief overview that follows, we shall occasionally lump together the different pathologies under the rubric of a general 'amnesic syndrome'. But we concentrate mostly on H.M. and equivalent cases, since these allow fairly direct comparison with the animal data at the anatomical level. We shall then use the animal data for an attempted functional dissection of amnesia.

The damage sustained by H.M. involves the loss of hippocampus plus amygdala plus adjacent cortex. The clinical material (with the possible exception of cases such as R.B., who appears to have bilateral damage limited to area CA1; Zola-Morgan *et al.* 1986) does not allow us to tell which structures are crucial for the amnesia, nor what specific contribution is made by the hippocampal formation. There is now a voluminous animal

literature investigating the different aspects of H.M.-type lesions, usually labelled H+A+ (for hippocampus + surrounding cortex plus amygdala + surrounding cortex). It has become clear that the contribution of the amygdala itself to the observed memory effects was greatly overestimated in the early literature, because of incidental damage to fibres of passage (see Murray 1996). Conversely, it has recently become evident that damage to perirhinal and parahippocampal cortex alone can produce major memory deficits (e.g. Zola-Morgan *et al.* 1989; Suzuki *et al.* 1993; see Murray 1996). In some memory tests at least, damage limited to the hippocampus proper may produce negligible effects (see Murray 1996); although this does not exclude a role for the hippocampus proper since, within the temporal lobe, the effects of damage appear to summate to a large extent across structures (e.g. Meunier *et al.* 1993; Zola-Morgan *et al.* 1994; but see Meunier *et al.* 1996). In sum, when considering the functions of the hippocampal formation or comparing animal hippocampal lesions with human clinical material, it must be remembered that even in cases in which the pathology is restricted to the temporal lobe, the full-blown amnesic syndrome appears to depend as much or more on damage to extra-hippocampal structures as on damage to the hippocampus proper.

The latter point will also be important when we come to consider the effects on memory of the anxiolytic drugs (Chapter 11, Section 11.1). A key fact for our theory is that the anxiolytics impair the control of theta rhythm (Appendix 5). From this we infer that the drugs should produce an impairment equivalent to that seen after modest damage to the hippocampal formation (including entorhinal cortex and posterior cingulate). But, at present, we have no evidence that theta is crucial for perirhinal or parahippocampal function. A lack of action on these structures, and/or only modest impairment of hippocampal formation function generally, would account for the relatively mild amnesic effects of the anxiolytics when these are given at the lower doses that are effective in treating clinical anxiety.

8.3 TYPES OF MEMORY

The third myth we must dispel is the idea that amnesia involves *loss* of memory.

In reviewing the clinical literature in the first edition we extracted three major principles. These rather global principles have stood the test of time. First, there is much that amnesic patients remember even with the full H+A+ lesion; furthermore what they remember is what they would be expected to remember, given the non-human data. Second, a powerful factor favouring the appearance of memory problems in an amnesic is the presence of sources of interference. A detailed analysis of this specific aspect of the data will provide a crucial link to the core of our theory. Third, there does not appear to be any disruption in consolidation as such. This is a contentious issue which we shall also subject to detailed analysis. From the point of view of the theory outlined in Chapter 5, these are all important points; but, despite them, we cannot too readily ignore the 'practically irresistible' clinical impression that these patients 'cannot form a durable record of their new experiences' (Weiskrantz 1978). We shall, therefore, not only criticize memorial views of hippocampal function, but also supply an explanation for the apparent loss of memories resulting from hippocampal lesions.

The emphasis on *amnesia* in this literature arises because of the relatively normal IQ and other intellectual functions of many such patients. For example, Jenny Ogden 'decided to assess H.M.'s ability to measure time and told him that he would be left alone for a period, and then asked to estimate the time that had elapsed. [She] left the room at 2.05 pm and returned at 2.17 pm. When H.M. was asked how many minutes had passed, he replied, without hesitation, "12 minutes; got you there!" There was a large clock on the wall and H.M. had noted the time when [Jenny Ogden] left, and continued to rehearse it while looking at the clock. When asked how many minutes had passed he simply subtracted the time he was rehearsing from the time shown on the clock. This anecdote illustrates more about H.M.'s sense of humour, his willingness to cooperate with anything experimenters may come up with, and his intact intellect, than about his ability to estimate time without memories!' (Ogden and Corkin 1991, p. 199). Superficially, then, H.M. appears to have a quite specific problem with memory. Yet this very anecdote shows that, at least by rehearsing (and rehearsing why he is rehearsing), he can under some conditions produce apparently intact memory performance.

Such signs of intact learning and memory in 'amnesics' have been seen under a wide diversity of conditions. The earliest reports concerned relatively simple motor tasks. Thus Corkin (1968) reported that H.M. showed good learning to track in the pursuit rotor and to tap out a sequential pattern. Sidman *et al.* (1968), also studying H.M., used operant conditioning techniques to establish good learning in a discrimination of circles from ellipses and in matching-to-sample with trigrams and ellipses. Similarly, Weiskrantz and Warrington (1979) reported successful eye-blink conditioning in two general amnesics, with, as in Sidman *et al.*'s (1968) experiment, retention lasting over a 24-hour interval. As reviewed recently by Ogden and Corkin (1991), H.M.'s intact memorial functions include immediate memory (e.g. digit span), remote memory (factual information learned before he was 17 years old), and implicit memory (e.g. the simple motor tasks described above). In contrast, his long-term, explicit memories from 17 years old onwards are largely absent (the exceptions seem to be items that have been for one reason or another extensively rehearsed). H.M.'s explicit memory is faulty for both semantic (general knowledge) and episodic (specific experience-related) items. His memory for verbal and non-verbal items is equally bad, which can be attributed to the bilaterality of his lesions.[1] It is consistent with our theory that H.M. also appears unusually lacking in anxiety (J. Ogden, personal communication)—but this observation can hardly be construed as proof !

It is this contrast between the 'amnesia', on the one hand, and the demonstrations of normal learning and memory in specific tasks, on the other, that has generated a large literature devoted to the analysis of 'types' of memory which are 'hippocampal-dependent' or not. There is no question that this idea has had considerable pragmatic utility for clinical assessment (see Butters and Delis 1995). However, it has usually carried the implication that the hippocampus stores (or helps to store) one categorical type of memory, while some other brain area stores another type (for recent overviews of some of the critical issues involved in these distinctions, see Baddeley 1988 and Chapter 2 in

1. Verbal memory is usually affected more by lesions of the hippocampus/amygdala of the language-dominant hemisphere, non-verbal more by similar lesions of the opposite hemisphere, with bilateral damage producing more severe effects than the sum of the unilateral effects. It also appears that 'interpretation of language involves widespread distributed systems bilaterally' (Bottini *et al.* 1994, p. 1241).

Fuster 1995). Squire and Zola-Morgan (1984) have tabulated the following variants of the suggested dichotomy (where in each case the first of the pair is the hippocampally-dependent one): fact/skill; declarative/procedural; memory/habit; knowing that/knowing how; locale/taxon; cognitive mediation/semantic; conscious recollection/skills; elaboration/integration; memory with record/memory without record; autobiographical/perceptual; representational/dispositional. To which can be added: declarative/non-declarative (Squire 1992; Robbins 1996); working/reference (Olton 1978a); conscious/automatic; instance/rule (Shanks and St. John 1994); data based/expectancy based (Kesner and Beers 1988); conceptual/perceptual (see Gabrieli 1995); contextual/associative (Hirsh 1974; Miller 1991); explicit/implicit (Baddeley and Wilson 1994); intermediate/long term (Rawlins 1985); episodic/semantic (Tulving 1984); relational/associative (Cohen and Eichenbaum 1993); configural/associative (Rudy and Sutherland 1989); and probably a few more! It is strange that a single area of literature should have spawned so many different alternatives (in detail, not just in terminology), the vast majority still current. But we think there is a simple explanation. What we are seeing is an attempt to describe, in terms of dichotomies of memory type, a continuum which is non-memorial. However, the apparently fundamental intractability of the 'type' analysis is not the only reason for doubting its validity.

We shall look at some of these specific ideas in greater detail below. But we believe (as do others) that this general approach is fundamentally flawed; that the hippocampus does not store even intermediate, far less long-term, memories (although it may store 'program-like' information which alters its own processing); and that memory is distributed throughout the brain. On this last point, we agree with Fuster (1995, pp. 1–5) that:

> Memory is a functional property, among others, of each and all of the areas of the cerebral cortex, and thus of all cortical systems. This cardinal cognitive function is inherent in the fabric of the entire cortex and cannot be ascribed exclusively to any of its parts. . . . Because of the multiplicity of actual and potential connections in the cortex, each neuronal ensemble may be part of multiple networks and, thus, part of multiple representations. What most critically defines a network (i.e. a memory representation) is the ensemble of connections that has formed it. Relationship is of the essence and, in this sense, *all memory is associative*. . . . Association is an attribute of all memories, at the root of their genesis as well as their evocation. . . . [As a result] perceptual [declarative, episodic, and semantic] memory is mainly represented in posterior cortex. . . . Motor memory consists of representations of motor action in all its forms, from skeletal movement to the spoken language. It too is acquired and evoked through the senses but, once acquired, it is largely represented in the neocortex of the frontal lobe, which comprises roughly one third of the human neocortex. The most automatic and firmly established aspects of motor memory are represented outside the neocortex, notably in the basal ganglia and the cerebellum.

To this we can add affective memory in the amygdala (Chapter 6) and, by implication, the potential for memory (but not necessarily conscious retrieval) in all subcortical areas.

8.4 THE EFFECTS OF DELAY

The fourth myth we must dispel is the idea that amnesia necessarily involves increased forgetting.

In matching-to-sample tasks, H.M.'s performance deteriorated badly when a delay was introduced. In delayed matching of ellipses, his deficit was severe at delays greater than 24 seconds; with trigrams, his retention was normal for as long as 40 seconds. In the latter task, it was noted that he achieved good performance at this longer delay only with the aid of constant verbal rehearsal (Sidman *et al.* 1968). Matching based on ellipses and similar visual stimuli, by contrast, usually eliminates the possibility of verbal rehearsal.

Verbal rehearsal is particularly significant in the context of White's (1985) theory of direct remembering. In his view, stimuli in memory have similar properties to stimuli in a visual field, 'memory decay' being the direct result of the temporal distance at which the item is 'viewed'. Verbal rehearsal, while it can be continued, provides a means of retaining at least some aspects of the original information (those which can be verbally encoded), while by-passing the 'viewing' of the richer memory stimulus itself. White's model predicts that provided (a) strategies such as rehearsal are eliminated and (b) response bias is eliminated by signal detection analysis, then memory decay will follow a single exponential function. In particular, the rate of decay will be the same independent of the point in time at which it is estimated. His model accounts for over 95 per cent of the variance in memory decay in experiments with, for example, pigeons performing delayed matching-to-sample and rats performing delayed conditional discrimination (Kirk *et al.* 1988). It should be noted that, in the experiment with rats, care had to be taken to prevent the use of positional strategies as a form of rehearsal. We have found that, when undergraduate subjects perform delayed matching-to-sample with swastika-like stimuli and random checker boards, as with rats allowed to adopt positional strategies, there can be virtually no memory decay. In the checker board case, this is true even when the difficulty of the discrimination is increased to the point where overall performance becomes very poor. However, when verbal rehearsal is prevented by adding a competing verbal task in the delay interval (R. Kirk and N. McNaughton, unpublished data) or by using a circle size discrimination which is not amenable to verbal encoding (Money *et al.* 1992), then the expected exponential decay is observed. What is particularly important about this result is that the exponential function allows estimation of the discriminability of the positive versus the negative stimulus at a nominal zero delay (note that this is *not* the same as the discriminability obtained in a simultaneous matching to sample task). When amnesics, diagnosed as SDAT, are tested on this task they show perfectly normal memory decay rates, but their capacity to discriminate the alternatives at nominal time zero is greatly reduced (Fig. 8.1). This result is consistent with the fact that, if the level of initial learning is equated, rates of forgetting in amnesics do not appear different from controls (e.g. Freed *et al.* 1987).

We have the interesting situation, then, that amnesics can carry out matching-to-sample provided stimuli are maintained in iconic or similar storage, or provided any delay is bridged by rehearsal or a similar strategy, but not if the delay must be bridged by 'direct remembering' (White 1985). However, at least in the case of visual delayed matching-to-sample, faulty direct remembering in amnesics is not the result of an increased rate of memory decay (even though such an increase in decay *can* be produced by a variety of procedural and pharmacological manipulations). Rather, it is the result of a confusion of stimulus alternatives, a confusion that is present even at the very smallest

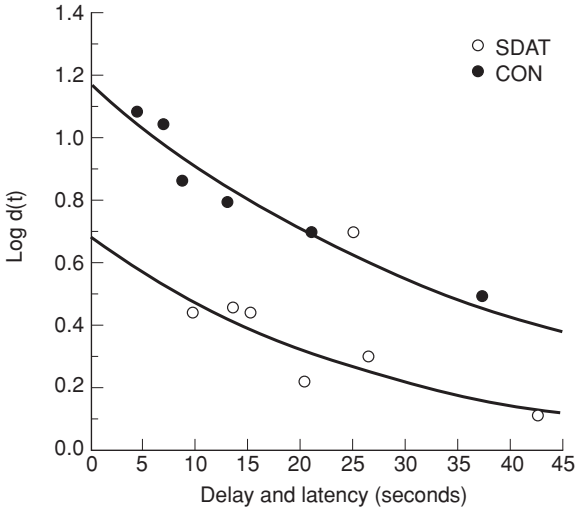


Fig. 8.1 Performance of patients diagnosed with senile dementia of the Alzheimer type (SDAT) and age-matched controls (CON) on a delayed matching-to-sample task. Performance was assessed using a bias-free measure (log *d*) derived from signal detection theory and plotted as a function of the actual delay between stimulus presentation and response (i.e. experimenter-imposed delay plus subject-imposed latency to respond after presentation of the choice stimuli). The curves show simple exponential decay functions fitted to these data (see text). SDAT showed a clear decrease in the parameter log *d* at nominal zero delay but no decrease in the parameter *b*, i.e. the rate of memory decay. (From Money *et al.* 1992.)

delays (e.g. 0.1 s). This is also the pattern of results seen in rats treated with either the anticholinergic scopolamine (Kirk *et al.* 1988) or the anxiolytic chlordiazepoxide (Tan *et al.* 1996). We comment on these findings further below. Thus, even when we observe a deficit with amnesics in the delayed (hence presumably memory) form of a task but not the non-delayed form, this deficit need not be attributed to a failure of memory in the conventional sense, but could be due to a fault in encoding, retrieval, or related processes. (We cannot be more definite, as we are attempting to generalize from SDAT patients to amnesics in general.)

There has been considerable debate in the animal literature as to whether, in fact, hippocampal lesions produce an increased rate of decay of memory. In all the experiments which have reported such an increase, there are one or more of four potential problems. First, per cent correct may be used as the measure of performance (see below). Second, and not usually recognized, there may be delay-related response biases (Herremans *et al.* 1995) which contaminate the data. Third, control values at shorter delays may be at ceiling (thus suppressing the capacity to show a difference due to hippocampal lesions). Fourth, and more contentiously since not everyone would accept White's exponential decay model, virtually none of the published data has attempted to fit that model. Since, in our own experiments mentioned above, departures from the model could always be traced to a failure to exclude species-specific, non-memorial strategies, this is likely to be an important omission.

The problems with per cent correct as a measure of performance have been adumbrated by Ringo (1988, 1991, 1993). He reanalysed a number of experiments in the literature using a *d'* or arcsine transformation to correct for the scaling problems of per cent correct and concluded that there is no evidence for an increased rate of forgetting. However, he ignores three points which need to be taken into account if this issue is to be settled. First, he failed to fit an exponential decay function and, when this is done to the data which he analysed, a rate change is observed (White and Harper 1996). Second, his reanalysis was by transformation of group means; strictly, it is the individual datum which should be transformed. Third, his *d'* calculation assumes that there is no response bias—a strange assumption, since *d'* was originally created so as to estimate discriminability independent of response biases when these are present. In this, as we shall see, he seems to have led some others astray.

Ringo's failure to emphasize the proper calculation of *d'* is important. Both *d'* and log *d* (the variant preferred by White) are bias-free measures (but see Herremans *et al.* 1995) based on the pattern of hits, misses, correct rejections, and false alarms. Per cent correct (among its other problems) collapses this signal detection matrix, losing the critical bias information. It is quite likely that interference, and hence bias, will increase with increasing intervals (Revusky 1971). It should also be noted that, when responding is close to the 100 per cent correct ceiling, neither *d'* nor log *d* can be properly estimated and a transformation of the mean *d'* is technically meaningless.

Our next concern is one shared by Ringo and is crucial whatever one's theories of the appropriate measures to calculate. Task difficulty must be set so that performance in the control group does not approach 100 per cent (and this must be done in ways which do not increase response bias or other confounding factors). In many of the published data on this issue (see Squire and Zola-Morgan 1984), per cent correct at short delays is close to 90 per cent in the hippocampally lesioned groups. With these values, it is virtually impossible to demonstrate any difference there might be from controls, since both groups are performing at ceiling. (It is worth noting that White's model can, in principle, circumvent this problem without requiring a change in procedure, since his functions can be estimated from any portion of the decay curve and then extrapolated back to obtain log *d* at zero delay.)

Finally, there is the issue of the exponential decay model. As noted above, this can be fit, provided suitable precautions are taken, across species (pigeons, rats, human beings) and across tasks (delayed matching-to-sample; delayed conditional discrimination). We cannot prove that it will always be applicable, but those involved in this debate should look very closely for non-memorial strategies (e.g. coding the to-be-remembered response by holding still in one position; verbal rehearsal) within any experiments the data from which do not fit the model. It is also essential to ensure that there are no ceiling or floor effects preventing detection of the exponential decay function.

These problems are exemplified by a recent study (Alvarez *et al.* 1994) which purported to show an increased rate of decay using *d'* as the measure of performance. First, it should be noted that the group x delay interaction in this experiment was non-significant when Ringo's *d'*-type transform was used. Second, the experimenters clearly did not calculate *d'* in the conventional fashion, as their data are such that this would have been impossible at the shortest delay. Third, the performance in both the control

and lesion groups was over 80 per cent correct in two out of the three delays, and the initial parts of the resultant curves are very far from exponential decay. Finally, the 'mixed delay' testing used in this experiment involved each delay being presented continuously for 3 days at a time. This clearly allows animals to adopt specific response strategies at specific delays, since the animal knows in advance which delay is operative. It is only when the different delays are interleaved that the hypotheses can be tested properly. An interesting comparison can be made with Alvarez *et al.* (1995), in which, between the 10- and 40-minute delays, controls and animals with hippocampal lesions showed virtually identical rates of decay.

Overall, then, there are no data at present to show that human amnesics or animals with H+A+ lesions have increased rates of forgetting, and some evidence suggests that they do not. However, the matter must be seen as open, and potentially complex. When proper signal-detection measures are submitted to exponential analysis, impaired memory with no increased forgetting has been seen, as noted above, with both SDAT patients and rats treated with anticholinergics or benzodiazepines; however, increased rates of forgetting with little deficit at zero delay have been observed with medial septal lesions and with injections of phenobarbitone (Harper *et al.* 1994; see White *et al.* 1996, for review). If these results represent fractionations of the hippocampal system deficit, then hippocampal amnesia may involve both delay-dependent and delay-independent processes.

8.5 THE ROLE OF CONTEXT

The fifth myth we must dispel is that conditioning to the context of a stimulus plays a crucial role in amnesia.

'Context' and its consequences need to be defined carefully. Strictly, a context can only be said to exist when there is some separate identifiable stimulus also present with which the 'contextual' stimuli are linked in some way. Even in this form, the idea can require considerable qualification (Miller 1991, pp. 1–6). However, there are those who, starting with the idea that a box in which a CS–UCS pairing occurs is the context for that pairing, would then describe the box itself as a 'context' even when there is no embedded CS–UCS pairing. But it should be obvious that pairing a box with a shock is itself simple CS–UCS associative conditioning, but with a complex, spatially and temporally extended CS. To call this a 'context' is a plain misuse of language.

Although impairment by hippocampal lesions of conditioning of fear simply to a box has occasionally been reported, Phillips and LeDoux (1994) provided some comfort to a contextual view of hippocampal function in an experiment involving three conditions: 'forward pairing of a phasic tone conditioned stimulus (CS) with a footshock unconditioned stimulus (US), unpaired presentations of the CS and US, or presentations of the US alone. All three procedures resulted in the acquisition of conditioned freezing to the box in which these manipulations took place. Lesions of the dorsal hippocampus prevented the acquisition of contextual conditioning in the paired procedure, as also reported previously, but not in the unpaired or US alone procedures.' Despite the authors' description of the box as a 'context' in all three cases, their results show that it is only when the box provides a true context for the CS–UCS conditioning that hippocampal lesions have an effect.

Note, however, that what Phillips and LeDoux showed is not that hippocampal lesions impair the use of a context in the CS–UCS conditioning but rather that, when there is the opportunity simultaneously to form both a discrete CS–UCS association and a non-discrete CS–UCS association, then hippocampal lesions selectively impair the latter. By contrast, when a context is explicitly used to enhance the effects of conditioning, hippocampal lesions have no effect (Gisquet-Verrier and Schenk 1994; Hall *et al.* 1996). Conversely, hippocampal lesions do impair fear conditioning to a simple diffuse stimulus—food deprivation—even when this cannot be viewed as a 'background' to some other conditioning (Davidson and Jarrard 1993). In the latter paper, it was shown that conditioning to a continuous auditory tone was intact. All of these data argue against a role for the hippocampus in conditioning to context, in the strict sense, and suggest that some interaction of the stimulus parameters (diffuse nature, temporal continuity, etc.) with the response system is important in making the tasks sensitive to hippocampal lesions. (Stimulus type alone is clearly not sufficient, as rats with hippocampal lesions can learn a conditional drug state–flavour aversion discrimination, and drug state would appear to be as diffuse a stimulus as hunger state; Skinner *et al.* 1994.) We consider this issue further below.

In a further variation on the theme of context, Winocur (1981) has proposed that, under conditions of high interference, animals with hippocampal lesions are unable (for whatever reason) to extract enough information from cues that are typically sufficient for controls. In consequence, they become abnormally dependent on 'contextual cues', i.e. background stimuli that can be used to determine which of several possible responses to a common stimulus is correct in a particular context.

In support of this view, Winocur and Olds (1978) trained rats on a simultaneous discrimination between horizontal and vertical stripes and then tested them after a 7-day retention interval either in the same or a different apparatus, and either under the same (food) or a different (water) drive. There was a retention deficit only when conditions were changed between acquisition and retention testing. When reversal learning was substituted for retention testing, the opposite occurred: there was a large deficit in the unchanged conditions, and only a non-significant one in the changed conditions. Thus, in accord with Winocur's hypothesis, the lesioned animals were excessively dependent on contextual cues for the maintenance of the behaviour which they had first learned. (Note that this result is opposite to that predicted if hippocampal lesions *impair* conditioning to context.)

A rather different suggestion, also concerning context, has been made by Hirsh (1974). According to this proposal, hippocampal lesions disrupt the 'contextual' retrieval of information from memory. Since this proposal has been somewhat neglected, and is rather technically defined, we shall go into it in some detail, largely in Hirsh's own words.

> Contextual retrieval is defined as the retrieval of an item of stored information initiated by a cue which refers to but is not necessarily described within the information that is retrieved. An example is the indexing of library books according to their authors. This concept has strong implications for theories of how the brain controls behaviour. Understanding these implications requires defining and exploring the metaphysiological concept of the performance line. A performance line is defined as a system mediating the series of events or processes initiated by the overtly observable stimulus

and resulting in the occurrence of the overtly observable response. It is considered to exist in real time and real space and ultimately to be physiologically observable.

S–R theories, as usually interpreted by physiological psychologists, hold that memory is stored upon the performance line. The stimulus is defined as activation by an environmental event of a neural system sensitive to it. Memory, or more exactly learning, is held to result from the formation of a functional connection between the neural elements sensitive to the stimulus and those responsible for producing the response. This connection becomes the key part of the performance line for that particular combination of stimulus and response elements and is unique to it. . . .

When contextual retrieval is present it is no longer necessary to store acquired information upon the performance line. Information can be stored in a sequestered place or state free from interference by information processing being carried out on the performance line. Schematic representations of systems with and without contextual retrieval are presented in Figure [8.2]. In a system with contextual retrieval, information can be moved from storage to the performance line independently, and thus before the occurrence, of the stimulus described within the retrieved information. . . .

It is assumed that when an item of information is contextually retrieved it is placed upon the performance line. Perceptual systems are adjusted to be particularly receptive to the stimulus described within the information. The object of attention is specified. In such a system behaviour is no longer completely controlled by environmental events because the perceptual systems are adjusted prior to the occurrence of the overtly observable stimulus. The potential of the retrieved stimulus description for controlling behaviour is considerably greater than that of non-retrieved descriptions. . . . By contrast, when memory is stored upon the performance line, no stimulus 'description' has more potential for controlling behaviour than any other until the event described actually occurs and initiates retrieval. . . .

It is also assumed that motor systems are adjusted to more readily produce the response described within the retrieved information as a result of contextual retrieval . . .

When behaviour which conflicts with previous learning is acquired, the previous acquisition must be made inoperative when learning involves altering the performance line. The new and previous acquisitions cannot both be retained, as both would be activated by the stimulus. To avoid conflict at least part of the previously acquired learning must be removed from the performance line, usually by inhibiting one of its stages.

Systems utilizing contextual retrieval do not require deletion of previous learning. The conflicting items of information can be differentiated by the addition of a contextual label indicating that the previously acquired information was formerly true. . . .

The effects of hippocampal lesions are held to result from the disruption of contextual retrieval. In the normal brain, storage is considered to be separate from the performance line. When the hippocampus is absent storage is held to occur within the performance line. (Hirsh 1974, pp. 422–6 [our figure reference].)

Hirsh applies this theory to the whole gamut of behavioural changes seen after hippocampal lesions. But, since we are concerned here with purely memorial approaches to hippocampal function, we note only that the general theory of septo-hippocampal function which we present has much in common with Hirsh's ideas, except that we would see the ordering of contextually retrieved goals as being effected by the prefrontal cortex rather than the hippocampus, with the hippocampus becoming involved only when there is significant conflict between goals at a specific point in time. As applied to the particular problem of memory, Hirsh's theory treats the hippocampal deficit (when it occurs) as arising from an inability to use contextual labelling to decide between

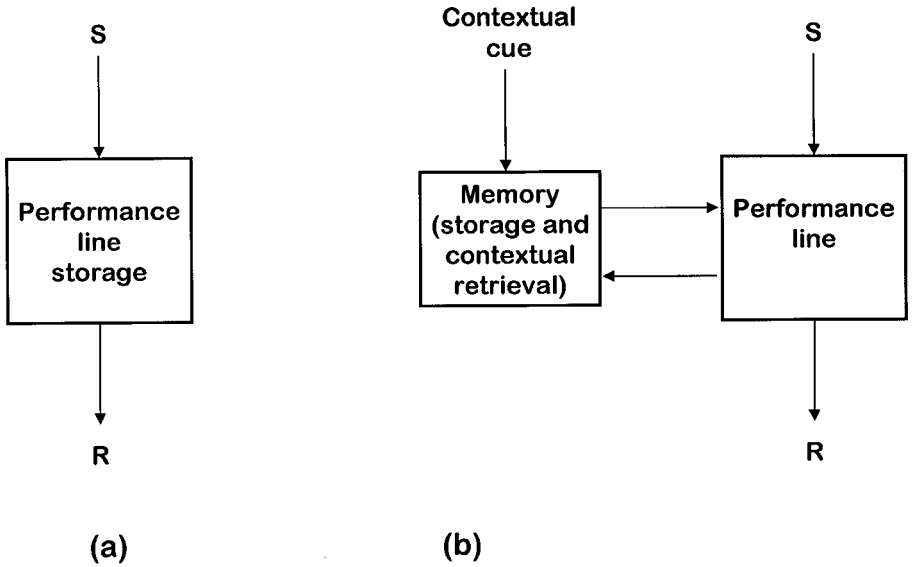


Fig. 8.2 Schematic diagrams of systems using associative (a) and contextual (b) retrieval. S, stimulus; R, response. (Redrawn from Hirsh 1974.)

competing response tendencies under conditions of high interference. Note that the requirement for the presence of interference distinguishes this view from the simple contextual views discussed above. We shall consider this aspect in much greater detail when we give our own account.

As far as the concept of context goes, Hirsh's (1974) view of its role in the behaviour of animals with hippocampal lesions is apparently very different from Winocur's (1981). Indeed, Winocur himself (personal communication) sees the two approaches as diametrically opposed. But a little reflection shows that these two workers use the term 'context' in different ways. For Hirsh, this is something the animal puts into the description of the stimuli to which it responds; for Winocur, it is out in the environment. The process that Hirsh ascribes to the hippocampus would play a particularly critical role in the control of behaviour when there are no large changes in the environmental context (Winocur) to assume this control. Thus both Hirsh and Winocur should predict that a large difference in the environmental context between initial acquisition and reversal training would help the animal to overcome its usual deficit in reversal learning, as was in fact found by Winocur and Olds (1978). So, from this point of view, rather than being opposed to each other, the two hypotheses are mutually complementary. Hirsh's mechanism can be seen as specifying the reasons why, as postulated by Winocur, animals with hippocampal lesions are unable to extract enough information from cues that are typically sufficient for controls: they lack the process of contextual labelling and retrieval. In line with this suggestion, Brown (1982) describes single units in the hippocampus which, in rhesus monkeys trained on a conditional discrimination, responded, or failed to respond, to particular stimuli depending on the context in which these were presented.

As we have noted, Hirsh's views are very close to our own (and quite close to the relational view of Cohen and Eichenbaum 1993); but, as with a number of other theories of hippocampal function with which we contrast our position, we believe that his argument should be turned exactly upside down. In our view, hippocampal lesions do not so much cause the animal to fail to label a particular stimulus with its context, but rather permit too many stimuli to be so labelled. Contextual errors occur, then, not because the key item is not retrieved by the context, but because many other items are concurrently retrieved and the correct item is then lost in the pandemonium. On this view, contextual errors should be 'intrusion' errors: the retrieval of information associated with the context but which is currently incorrect. On a pure contextual labelling point of view, errors should be random. The same arguments apply to Miller's (1991) theory, which sees the hippocampal theta rhythm as indexing contextual information in the brain.

Probably the most explicit case of contextual control of retrieval is conditional discrimination. This is usually impaired in animals by hippocampal lesions, as it is also in human beings with temporal lobe lesions (Daum *et al.* 1991). The dissociation of intact non-conditional and disrupted conditional discrimination is clear even with simple classically conditioned responses (Ross *et al.* 1984; Daum *et al.* 1991). One interesting failure, however, to produce a deficit of this kind is that of Whishaw and Tomie (1991), who used a string-pulling task in which size of string predicted the position of reward. Observation of the rats suggested that they were not turning the task into a simple go–no-go discrimination by adopting a position bias, and that some feature of the stimuli (such as their contiguity to the rat) may have removed the normal deficit through the loss of some source of interference. Similarly, with delayed conditional discrimination, the size of the deficit appears to depend on the extent to which there is interference (Wan *et al.* 1994). All of this is consistent with Hirsh's general position.

The most recent development of this general contextual orientation is Gaffan's analysis (e.g. 1994) of 'scene specific memory for objects'. He used computer-generated 'objects' and 'backgrounds' to create 'scenes'. This allowed careful variation of the precise relationships of objects to backgrounds and, crucially, permitted testing of monkeys after learning of many hundreds of discriminations against a multitude of backgrounds. Thus Gaffan was able to test the monkeys on specific tasks again after experience approximating the normal human situation. As noted above, the monkeys in these experiments showed the same deficits as human beings when this experiential background was matched. More relevant to the present argument was Gaffan's manipulation of the role of context (in his terms background or scene specificity) in conditioning. 'The animals learned four types of discrimination problem: (1) object-in-place discrimination learning, in which the correct (rewarded) response was to a particular object that always occupied the same place in a particular unique background, (2) place discrimination learning, in which the correct response was to a particular place in a unique background, with no distinctive object in that place, (3) object discrimination learning in unique backgrounds, in which the correct response was to a particular object that could occupy one or the other of two possible places in a unique background, and (4) object discrimination learning in varying backgrounds, in which the correct response was to a particular object that could appear at any place in any background. The severest impairment produced by fornix

transection was in object-in-place learning. Fornix transection did not impair object discrimination learning in varying backgrounds. The results from the other two types of learning task showed intermediate severity of impairment in the fornix-transected animals' (Gaffan 1994, p. 305).

Gaffan argues that the varying background case makes background irrelevant and so eliminates interference from previous background-related learning. In the object-in-background case, the control subjects learned particularly well. Gaffan argues that this is because they can use the specific background and spatial contextual information to eliminate the interference from previous background-related learning; and that this active removal of interference is prevented by the fornix lesions. This is a specific form of a more general hypothesis which we advance below. Gaffan also suggests that 'if different features of a complex scene are represented in the various cortical areas that are linked together via the hippocampus–fornix–mammillary system, then this system might connect those features together into a widespread cortical distributed representation of the scene' (Gaffan 1994, p. 314). Here, as with our treatment of other theorists, we shall invert this proposal and suggest that the hippocampus *prevents* connection of features which would otherwise confuse the distributed representation of the scene.

8.6 A PRELIMINARY THEORY OF HIPPOCAMPAL AMNESIA

So far we have largely concentrated on issues which cut across all of the current theories of hippocampal function. At a number of points, however, we have explicitly or implicitly dealt with those current theories that are particularly tightly tied to such issues (e.g. Hirsh's views on context). In what follows we consider a selection of the various theories of what distinguishes hippocampal-dependent from hippocampal-independent memory, together with the data which impinge most directly on each theory. At various points we provide expositions of the theories; but it is an interesting aspect of this area of research that many views of hippocampal function are based on the results from just one or two key experiments and can be expounded in a sentence or two, since they are only sparsely detailed in terms of learning theory, cognitive psychology, neuroanatomy, or neurophysiology. As above, we shall restrict this analysis to the large literature on gross lesions to the hippocampal formation and related structures (in essence H+A+ lesions). (We address the issue of the differential effects of selective lesions to specific parts of the septo-hippocampal system in Appendix 9.)

Our own view requires quite detailed specification, which it receives in the next two chapters. But, for present purposes, it can be summarized as follows. The hippocampus detects and eliminates conflict between nearly equally primed incompatible goals (see Fig. 1.7). These goals may be distinguished both by distinct response tendencies (e.g. approach–avoidance) and by distinct stimuli to which a response could be made (e.g. left lever–right lever). Memory experiments will usually involve the latter. The detection of conflict often results in information-seeking behaviour (e.g. risk assessment), but can also result in internal reassessment of older information via an increase in the valence of affectively negative features or associations of stimuli. In memory experiments risk assessment

behaviour is likely to be minimal, as measured by standard tests. Under these circumstances, the most obvious result of output from the behavioural inhibition system is that of behavioural inhibition itself, in the form of suppression of any prepotent tendencies to respond to incorrect alternatives. In experiments involving threatening stimuli, increased attention to such stimuli and increased negative valence of the associations to them on future trials give rise to anxiety.

This hypothesis is intended to apply equally to all species. That does not mean that all species will show the same sensitivities to hippocampal lesions in the face of any specific experimental paradigm, since reliance on, for example, the verbal channel in human beings or on olfactory stimuli in rats will greatly modify how any particular problem is solved. The hypothesis implies that removal of the hippocampus does not cause a loss of memory as such (as opposed to failures to control memory or to show suitably modified behaviour). It implies also that the amnesia following hippocampal damage does not involve an increased rate of forgetting (except where increasing delay increases some source of conflict).

To get a further flavour of the implications of this hypothesis, let us look at the challenge to a truly memorial view of amnesia raised by the method of cued recall (Weiskrantz 1978). In this paradigm, the subjects are given partial information about an item they are to recall and asked to identify the complete item. In different versions of the paradigm this information may be fragments of a picture, or word, the first few letters of a word, a semantic clue to the item, or the first of a pair of words that rhyme. With repeated exposure of the incomplete cue followed by the complete item, subjects require less and less partial information to identify the item, or identify it more rapidly, thus providing a learning curve. Alternatively, the cue can be presented at the time of retention testing. Using these techniques, Warrington and Weiskrantz (e.g. 1970, 1974; Weiskrantz and Warrington 1970a,b) have shown that Korsakoff amnesics can learn as fast as normal subjects and can retain the information over hours or days. As already noted, Korsakoff patients do not present a good model of temporal lobe amnesia, but Milner *et al.* (1968) reported a similar demonstration using incomplete pictures with H.M. (see also Ogden and Corkin 1991, p. 200). The normal or nearly normal performance of the amnesic subjects is not simply due to the task's being easier, since they are disproportionately aided by the provision of cues (Warrington and Weiskrantz 1970, 1974).

As with our earlier discussion of the delay data, cued recall suggests that in amnesics the memory as such is intact, but that there is some problem with the capacity to store or retrieve it correctly. This further suggests that what characterizes amnesic-sensitive, and potentially also hippocampal-sensitive, tasks is not a particular type of memory at all. Rather, it is some specific task demand, unrelated to memory *per se*, which makes these tasks sensitive to temporal lobe damage. The most obvious effect of partial cueing is to reduce the number of potentially correct answers and, as a mirror image of this argument, our theory is that the hippocampal formation normally functions, in the absence of partial cueing, to eliminate the incorrect alternatives. Warrington and Weiskrantz (1978) make a specific suggestion as to the task demands which give rise to this situation, but this will be best dealt with at the end of the chapter, after we have considered various specific hypotheses arising from the non-human data (and tested in both human and non-human subjects).

8.7 RELATIONAL MEMORY

A good way to illustrate our own view is to compare it directly with one or more of the current theories. We devote separate sections to a number of these below, but here we look in some detail at what is probably the most recent synthesis of the animal and human literature on the hippocampus and memory (Cohen and Eichenbaum 1993; Eichenbaum *et al.* 1994), as an exemplar of memory views in general. We have chosen to highlight this theory not only because it is one of the most recent, but also because it has been expounded at book length and we are more in sympathy with it than with many other particular views of the role of the hippocampus in memory. Furthermore, it is easy to see how their theory derives other current theories as special cases. We shall then describe the logical inversion which characterizes the fundamentally different approach of the present theory, before proceeding to a summary of the relationship of a number of current theories both to the present approach and to the data.

Eichenbaum *et al.* (1994; Cohen and Eichenbaum 1994) view the hippocampus as being critical for what they term relational memory. Key features of this proposal, with which we agree, are: that the hippocampus is not concerned with very short- or very long-term memory storage, these being functions of cortical areas; that the hippocampus is not concerned with simple associative learning (of the type we ascribed in Chapter 6 to, for example, long-term potentiation of sensory input to the amygdala); that the entorhinal cortex carries multimodal information to the hippocampus; and that an intact hippocampus is necessary if an animal is to learn many tasks (Eichenbaum *et al.* would say all tasks) that require relational processing.

However, as we noted in a commentary on the Eichenbaum *et al.* paper (McNaughton 1994), we believe that their theory (along with others) is flawed because of the specifically memorial perspective they take to the data. They deliberately and quite explicitly excluded from consideration many of the data on behavioural inhibition which have been central to our own approach. They also excluded all consideration of septal and monoaminergic input to the hippocampus and of subcortical output (a surprising omission since this is a major output in neuronal terms). Furthermore, in common with several of the more recent memory theories, they view the hippocampus as a form of intermediate store and do not make it at all clear (although they at least try; see Cohen and Eichenbaum 1993, pp. 68–72) how information stored in the hippocampus can be transferred to distributed networks (and distributed they must be, given the complex polymodal associations usually involved) in the neocortex.

In what follows, the following points should be borne in mind. First, as argued in Appendix 7, long-term potentiation in the hippocampus is much more likely to be a general attention-like, or programming-like, mechanism than it is to be the basis of detailed storage of complex sensory information. Second, it is difficult to reconcile a role for the hippocampus as an active creator of memory with its involvement in innate responses (e.g. innate conflict) and with the contrast between runway acquisition (intact after hippocampal lesions) and extinction (impaired). Third, for many situations, if the hippocampus were to inhibit the formation or the recall of incorrect memories, this would produce similar (but not identical) results to its actively forming correct ones. (The differences occur in subtle details, such as the fact that errors made by amnesics

are frequently intrusions of items which would have been correct on previous trials, a fact not dealt with by the relational theory.)

With these points in mind let us now consider what Cohen and Eichenbaum (1993) mean by relational memory, how this encompasses some of the other current hypotheses of 'hippocampal memory', and the implied consequences of their analysis for our own position.

At root, their theory is based on the declarative–procedural distinction (knowing what as opposed to knowing how). They propose 'that the hippocampal system plays a critical role in mediating declarative memory, one of the memory systems of the brain' (Cohen and Eichenbaum 1993, p. 60). What is distinctive is their elaboration of the critical features of declarative memory. 'Input to the hippocampal system consists of highly processed data from many higher-order cortical processing areas, each representing *functional descriptors* of the perceptually distinct objects and events encountered during learning, plus information about the affective and behavioural responses elicited' (p. 61). A crucial point is that 'representation of scenes, events, and complex ideas in our proposed declarative memory system is *not* as blends . . . and does *not* involve conjoining of the multiple individual stimuli or constituent pieces of knowledge into unified knowledge structures . . . [It] preserves the status of the constituents of the larger structure while still permitting the larger structure to be appreciated' (p. 64). Thus, declarative memory, within their theory, consists of a wide variety of informational entities (not just objects or qualities but also, for example, behavioural responses) all interconnected by the relations which each holds to the other: e.g. bird—(is an)—animal; canary—(is)—yellow. Storage of items belonging to the declarative system is presumed to be located within the neocortex.

A declarative memory system with these properties must suffer from a 'binding' problem. Within perception, the equivalent problem is: given the multiple unimodal systems, each dealing with a separate aspect of the visual world (e.g. colour), how do we obtain a unified picture of the world? How are the separate parts bound back together again? The declarative memory system (and a variety of other types of complex parallel associative memory systems) suffers from the memorial counterpart to this problem. With each node representing an entity and connections between nodes representing relations, how, at any one time, do we ensure that the right combination of nodes and connections is active? One potential solution to this problem is that there is an independent (hippocampal) indexing system (Teyler and DiScenna 1986). On this view the hippocampus stores (via long-term potentiation) the 'index of neocortical areas activated by experiential events' (Teyler and DiScenna 1986, p. 147). In this pure form, of course, this view fails because loss of the index should eliminate *all* declarative memory; yet, for example, H. M.'s longer-term memories are intact. However, in Cohen and Eichenbaum's version, 'the hippocampal system . . . apparently stores—for a time after learning—a representation from which declarative memory of the event or scene can be *reconstructed* from the permanent representations of the various attributes or features of the event or scene stored within the various neocortical processors' (Cohen and Eichenbaum 1993, p. 288).

This resolution appears rather neat until we notice that it does not solve the binding problem as such. If the neocortical networks can, through strengthening of connections,

ultimately represent the declarative memory, then, intrinsically, they have no need of a separate processor to produce the binding itself—all of the necessary bonds are potentially there already. Alternatively, if the hippocampus is necessary for binding as such, its removal should degrade all memories, whatever their age. Let us run this argument backwards. Let us suppose that all of the necessary connections and potential bindings are present *ab initio*. What will happen during the initial phases of learning (or of consolidation, which we can view as a form of internal rehearsal of a specific learning experience)? With items A, B, and C, all with potential relations just subthreshold, and with the particular desired relation (perhaps temporally ordered) A–B–C, how do we prevent the coactivation of A, B, and C from generating B–A–C or C–A–B or, even more likely, the higher-order structure where all possible binary and ternary links between A, B, and C are equally strengthened? The binding problem in memory, then, is not about how we take inherently separate units of information (as in perception) and ensure that they are bound together; for these units will be bound by simple associative coincidence. Rather, the problem is about how we take the multitude of potential relations (which must exist as subthreshold connections before the start of learning, if the relation is *ever* to be formed in the neocortex) and *prevent* excessive binding. In other words, how do we inhibit the formation of incorrect linkages, how do we prevent a superfluity of incorrect memories? We believe the theory of this book supplies the answer to this question.

We now move from these criticisms of the relational memory theory to criticism of a range of alternative memory theories. In many cases we shall find that these reduce to special cases of the relational theory (see Cohen and Eichenbaum 1993, Chapter 11) and hence will be subject to the same logical inversion we intend to apply to the latter. However, we shall also consider a number of specific points of detail at which our view can be compared with these other theories in ways which do not map completely onto the relational approach.

8.8 CONFIGURAL MEMORY

A 'configural' hypothesis of hippocampal function has been proposed by Rudy and Sutherland (1989; Sutherland and Rudy 1989), largely on the basis of a single experimental result: that hippocampectomized animals failed to solve a discrimination in which two stimuli were reinforced if presented separately (L+, T+, with L standing for 'light' and T for 'tone'), but were non-reinforced if presented together (LT–). They (Alvaro and Rudy 1995a) have also recently shown a similar failure on 'transverse patterning' (i.e. A+B–, B+C–, C+A–), which was dissociated from a lack of failure on a set of equivalent elemental problems (i.e. A+B–, C+D–, E+F–). From this they conclude that the hippocampus is necessary for configural associations, that is, for 'the acquisition of a representation of a compound stimulus that is distinct from its elements' (Rudy and Sutherland 1989, p. 97). As they note, their experimental result is inconsistent with a working memory view, any temporal buffer view, and any spatial view of hippocampal function. It is also difficult to assess in terms of the procedural–declarative distinction, since it is not clear whether the task studied is declarative, given that the simple associative version of the task with the same stimuli is not. They note, however, that their

results are consistent with Hirsh's contextual view, and to this we can add also the relational view.

Hippocampal lesions do not always produce an impairment in configural discrimination. In particular, Whishaw and Tomie (1991) showed that while some of their rats (all of which had massive hippocampal lesions) were impaired in learning a configural discrimination, others were not, the fastest learner of all being one with lesions. They also showed that with a lesion made after acquisition of the task, there was excellent post-operative retention. They point to a number of features of their procedure which might account for this: they used a string-pulling task which maximizes the contiguity of cues; they used a tactile–olfactory compound rather than a tone–light one; and the rats had a choice between a correct and an incorrect response (rather than the go–no-go of Rudy and Sutherland's experiment). Cue contiguity does not appear to be the crucial factor, as a visual–tactile version of the task is also not impaired (Whishaw and Tomie 1995). None of the variations, of course, changes the configural nature of the task from a theoretical standpoint. Other differences from the original study by Rudy and his collaborators, which Whishaw and Tomis do not mention, were that all of the stimuli used were compounds formed from one of two alternative tactile stimuli and one of two alternative olfactory stimuli, and that hence both positively and negatively reinforced compound stimuli were presented (rather than the L+, T+, LT– of the original experiment).

Cohen and Eichenbaum (1993, pp. 165–71, 260–3) have provided a detailed critique of the configural view. Their main points are that 'although Sutherland and Rudy have attempted to extend the proposal beyond the initial paradigms of interest to them, it remains a limited-domain account . . . [and] it is in the *absence* of the hippocampal system that animals fuse perceptually distinct stimuli into configurals, at least in some circumstances' (pp. 261–262). Davidson and Jarrard (1992, p. 90; but see also Rudy and Sutherland 1992) also point out that 'rats solve serial–compound discriminations because stimulus B comes to modulate the capacity of stimulus A to elicit a response [rather than by compound formation]. Hence, although it appears to be the basis for solving a task purportedly explained by Sutherland and Rudy's (1989) theory, [the actual mechanism of] conditioned modulation does not seem to be encompassed by either their configural or their simple association view.'

Even with respect to the original finding it is not clear that configural theory holds. Thus, Davidson *et al.* (1993) found that neither the kainate/colchicine hippocampal lesion used by Rudy and Sutherland (1989) nor ibotenic acid lesions affected the negative patterning task. Although they do not remark on this, it seems likely that the reason for the discrepancy is that Rudy and Sutherland used continuous reinforcement of the positive stimuli while Davidson *et al.* (1993) used what appears to have been a variable ratio (VR) 3 schedule. This increases rather than reduces the problem for configural theory posed by the negative result, but suggests that the deficit, when it occurs, may result from the degree of response competition, since, with continuous reinforcement, the secondary positive reinforcement accruing to the elements which make up the negative compound will be higher than with VR3 (see also Whishaw and Tomie 1991). Response competition is a particularly obvious explanation for the transverse patterning result (Alvardo and Rudy 1995a), and comparison over a number of configural tasks

which did or did not show lesion effects led Alvardo and Rudy (1995b, p. 183) to conclude 'that hippocampal formation involvement depends on the degree to which the elements of a compound (e.g. A and B) provide sources of generalized excitation that oppose the inhibitory control conditioned to a configural unit (AB)' rather than on configurality as such.

Overall, then, the configural theory does not account clearly for all the data to which it is intended to apply, and these encompass only a very limited selection of the available data on hippocampal lesions; in addition, the theory makes no attempt to encompass the known anatomy and physiology of the hippocampal formation.

8.9 SEMANTIC ENCODING

The next theory to be considered is derived from experiments with human subjects and can be applied to animals only with difficulty. Butters and Cermak (1975) suggest that amnesic subjects suffer from a disturbance in the process of 'semantic encoding'. Put like this, the encoding hypothesis is almost necessarily limited to the human case, since the notion of *semantic* encoding is closely tied to the details of experiments on verbal learning. It is supposed that, when a subject is presented with a word, this word can be committed to memory after being subjected to one or more of (at least) three different kinds of analysis: acoustic (what the word sounds like), associative (what it is associated with), or semantic (what it means). If semantic encoding were the chief deficit in the amnesic syndrome, the problem of relating the human to the non-human data would be acute indeed.

Much of the evidence on which the semantic encoding hypothesis rests was gathered in experiments in which Korsakoff patients were required to learn serial lists of words (Cermak and Butters 1972; Cermak *et al.* 1973; Butters and Cermak 1975). In the earliest of these reports, the list consisted of eight words, of which two were from each of four different semantic categories (animals, professions, etc.). Cueing by category name improved recall among controls, but not among Korsakoff patients. This is consistent with a failure of adequate semantic encoding at the time of learning in the patient group. Cermak *et al.* (1973) compared the effects on recall of rhyming and semantic cues respectively. Controls and amnesics were equally helped by rhyme, but the amnesics were less helped by category names. Associative cueing (e.g. 'table' as a cue for 'chair'), however, was as effective with amnesic subjects as with controls. Thus it appeared to be specifically semantic encoding that was deficient. A similar inference is suggested by an experiment in which the subjects were asked to detect repetitions in a long list of words. Besides genuine repetitions, the list contained some homonyms, synonyms, and associates. Korsakoff patients made as many correct identifications as controls, and no more false recognitions of either synonyms or neutral words, but they made more false recognitions of both homonyms and associates. Butters and Cermak (1975, p. 393) take these results to indicate that 'the Korsakoff patients were encoding the words on acoustic and associative dimensions, but were not encoding the semantic dimensions of the words to the extent that would allow the rejections of acoustically identical or highly associated words.'

Butters and Cermak (1975) pushed this analysis further by making use of the technique known as 'release from proactive interference'. Wickens (1970) had shown that, in normal subjects, proactive interference is reduced if the semantic category to which successive items belong is changed for a particular item in the list. Cermak *et al.* (1974) showed that this held true also for Korsakoff patients, provided that the change in category was relatively gross (from letter sets to number sets); but, unlike controls, the amnesics derived no benefit when the category shift was more subtle (from sets of animal names to a set of names of vegetables).

Butters and Cermak (1975) make a strong case, then, for a role for semantic encoding as an important parametric feature in making a task sensitive to the amnesic syndrome. Does this imply that a deficit in semantic encoding is the key feature of amnesia in human beings? As noted above, if this position is accepted, it reopens a substantial gulf between the human and animal data. Before we accept it, however, it is necessary to show that a deficit in semantic encoding can explain all the features of the amnesic syndrome, and also that there is not some more general principle, consistent with the non-human data, which could account for the apparent deficits in semantic encoding.

We shall shortly discuss in detail experiments in which cues are chosen specifically to prevent verbal rehearsal, and hence incidentally eliminate semantic encoding. In these, the use of stimulus differences which cannot be semantically encoded nonetheless allows an amnesic deficit to appear. There are also clear cases where amnesics are helped by semantic cueing (Winocur and Weiskrantz 1976; Winocur 1981), and the benefit they derive from this is not necessarily less than that derived by controls (Mayes *et al.*, cited by Baddeley 1981). The semantic encoding hypothesis thus fails to account for all the data on the human amnesic deficit.

Butters and Cermak (1975) themselves also offer an analysis of their results which allows them to be linked with non-human data. They suggest that 'the Korsakoff patients' encoding deficit may be related to a general impairment in their ability to attend to the relevant dimensions of stimuli' (Butters and Cermak 1975, p. 393). On this formulation, the deficit in semantic encoding is itself the manifestation of a wider difficulty, one that could in principle be found in other species.

There is some evidence for this proposal. Oscar-Berman and Samuels (1973, 1977) trained patients on a visual discrimination with stimuli differing simultaneously along four dimensions (form, colour, size, and position). After training was complete, transfer trials were conducted to determine which dimensions were controlling the subjects' choices. In comparison to other brain-damaged patients and to normal controls, the Korsakoff patients made use of fewer stimulus dimensions, concentrating especially on colour to the exclusion of other possibilities. This may be the same phenomenon as the one described in rats with hippocampal lesions by Samuels and Valian (1968), who found that the lesioned animals were poorer than controls at transferring from a simultaneous discrimination, in which brightness and position were perfectly correlated, to brightness alone. A failure in multidimensional stimulus analysis might also underlie the excessive dependence on environmental context described in human amnesics by Winocur and Kinsbourne (1978) and in rats with lesions of the hippocampus by Winocur and Olds (1978). In these experiments, change in a dominant environmental cue produced a retention deficit or, conversely, provision of such a cue eliminated the deficit. These

findings are explicable if we suppose that in human and non-human subjects alike, a general problem in multidimensional stimulus analysis occurs as a result of hippocampal damage and that this prevents subjects from learning about more subtle cues in their environment.

Note that none of the stimulus dimensions studied by Oscar-Berman and Samuels (1973) was more 'semantic' than the others. Thus, if there is a common thread linking their findings to those of Butters and Cermak (1975), it has to be an impairment in multidimensional stimulus analysis that underlies a deficit in semantic encoding rather than the other way round. We can, then, link even those cases where semantic coding appears important to more general processing deficits of a type which can account for hippocampal deficits also in non-human subjects.

8.10 RECOGNITION MEMORY

Gaffan (1972, 1974, 1977a,b) has proposed that hippocampal damage disrupts 'recognition' memory. As we shall see in the next section, this has much in common with Olton's (1978a; Olton *et al.* 1979a,b) 'working' memory theory.

Gaffan distinguishes between, on the one hand, recognition memory (or memory for familiarity) and, on the other, the associative memory that is responsible for turning stimuli into secondary reinforcers. He proposed that hippocampal lesions disrupt only the former. This hypothesis is in many respects similar to Vinogradova's (1975), according to which the hippocampus subserves orienting responses to novelty and their habituation as the stimulus becomes familiar.

In an initial experiment to test this hypothesis, Gaffan (1972) studied the effects of lesions of the fornix-fimbria on the rat's response to stimulus change in a T-maze, starting with one black and one white arm. Normal rats choose the changed arm, except when a frightening stimulus (loud noise) is added on the choice trial, when they choose the familiar arm. Animals with fornix lesions chose at random under both conditions. (Note the similarity to the effect of anxiolytic drugs described by Ison *et al.* 1966; see Appendix 1.) Gaffan attributes these findings to a failure on the part of rats with fornix lesions to discriminate the familiar from the changed arm.

In two further tasks, Gaffan (1974) used monkeys to delineate more clearly the distinction between recognition and associative memory. In one, the animal was initially presented with a set of 10 objects, one at a time, five of which were followed by food and five not. There then followed a retention test in which the monkey was faced with pairs of objects, of which one was constant and the other was one of the objects that had been presented initially. If the latter had initially been associated with food, the animal's task was to choose it; if it had not, the animal was required to choose the constant object. Thus, the animal was asked, 'Do you remember which objects were associated with food and which were not?' As predicted by Gaffan, fornix-lesioned animals were not different from controls on this task. The second task was similar, except that the animal was initially presented with only five objects, all followed by reward. These five, and five others, were each then paired with the constant object. If the non-constant object had previously been presented (and paired with reward, as in the first experiment)

the animal was required to choose it. If the non-constant object had not previously been presented (unlike the first experiment) the animals was required to choose the constant object. Thus it was asked, 'Do you remember seeing this object before?' Again, as predicted, fornix-lesioned monkeys were inferior to controls on this task. Furthermore, controls were better in the recognition task than in the test of associative memory, whereas the lesioned animals were better at association than recognition. This kind of task is very similar to many used to study human memory. A subject may be presented, for example, with a list of words or pictures and later asked whether items on a second list did or did not figure in the first. Gaffan (1977a) designed another experiment to come even closer to this model. The monkey was shown 25 pictures (on slides) per daily session, each picture occurring twice. Order of presentation was randomized but the same slides were used each day. The monkey's task was to press a panel only on the second appearance of any slide. This is analogous to asking a human subject to answer 'yes' or 'no' to the question 'have you seen this picture earlier today?'. Two rhesus monkeys achieved over 90 per cent accuracy even when an average of nine slides were presented between the first and second occurrence of any one slide. After fornix section, they were able to achieve accurate performance only up to an average of three intervening slides. Note that it is precisely on this kind of yes–no recognition task that the human amnesic impairment is particularly pronounced.

In another experiment, Gaffan (1977b) explicitly compared a spatial and a non-spatial version of list learning. In the non-spatial version, the monkey was successively shown three colours (transilluminated discs), followed three seconds later by one of six possible colours. Its task was to press one response panel if the test colour had been presented in the sample set of three, or another if it had not. The spatial version consisted of the successive illumination of three out of six discs, followed three seconds later by the illumination of one of them. The monkey's task was to press one panel if the test disc was in one of the positions first illuminated and another panel if it was not. Fornix-lesioned monkeys were impaired on the two tasks to the same extent and in the same manner: they made a great number of omission errors ('forgetting'), but only when the test stimulus was the first to have been presented in the initial list of three colours or positions. Both normal and lesioned monkeys showed a pronounced 'recency' effect, that is their recognition was better the more recently the test stimulus had appeared (i.e. the later in the list). These results show the operation of retroactive interference, which will be discussed later in this chapter.

There is some question as to how far human amnesics match the deficits found by Gaffan in rhesus monkeys. Brooks and Baddeley (cited by Baddeley 1981) repeated Gaffan's (1974) experiments on recognition and association memory using human subjects and pennies as rewards, but otherwise replicating his design. Unlike fornix-lesioned monkeys, amnesic human subjects were worse at associative than recognition memory. Given the problems in determining the nature of the damage in such patients discussed earlier, the monkey data should be given primacy; but it is probable that, as in other such cases, there is some subtle difference in the testing procedure or species-typical reactions which determines the difference in results. If this is the case, it would throw doubt on loss of recognition memory *per se* as being fundamental to hippocampal deficits.

The recognition view is to some extent a special case of the relational view (although it is not discussed by Cohen and Eichenbaum 1993), in that what distinguishes items is their relation in time (i.e. more or less recent). It has probably not found favour recently because of the relative difficulty it has in accounting for spatial (location of a submerged platform, impaired) versus non-spatial (location of a visible platform, unimpaired) performance in the water maze, and for the fact that deficits in delayed matching-to-sample (see section on working memory below) are obtained with trial-repetitive but not with trial-unique stimuli. In tasks of the latter kind, recognition as such is held constant, so the greater difficulty of trial-repetitive stimuli must lie in a factor separate from recognition. It should also be noted that hippocampal-lesioned rats can make recency discriminations (inherently more taxing in recognition terms than pure familiarity), provided the items as a whole are fairly recent and there are not too many intervening items (Fig. 3 in Chiba *et al.* 1994). Note also that like the other theories we have considered so far, the recognition hypothesis has trouble with simple tasks like extinction and DRL, is intended to apply only to the memory data, and makes no attempt to provide a theory in terms of the known anatomy and physiology of the hippocampal system.

8.11 WORKING MEMORY

Olton (1978a; Olton *et al.* 1979a,b) suggested that hippocampal lesions disrupt working memory. A task may require for its solution information which is constant from trial to trial or moment to moment, this being deemed to be held in 'reference' memory. 'Working' memory, by contrast, holds information which changes from trial to trial or moment to moment. The nature and function of working memory has received quite extensive analysis by cognitive psychologists over many years (see Baddeley 1986, for review). The paradigm example of a test of working memory is the delayed matching-to-sample task. Here, the animal is presented with a sample stimulus and then, after a delay, must choose it from a pair of stimuli. From trial to trial (each trial consisting of the sample presentation followed by the choice test) the sample stimulus is changed (the stimuli can be trial-unique, or the sample can be randomly chosen from a fixed pair of stimuli, or some variant between these two extremes), so that the animal can only make the correct choice by remembering the specific sample presented at the start of the current trial. This job is done by working memory, and is attributed by Olton *et al.* (1979a) to the hippocampus. Conversely, according to their proposal, the hippocampus is not concerned with memory of those features of the matching to sample task which are constant from trial to trial (e.g. the nature of the reward, the required motor response, the matching to sample rule itself).

A similar evolution in experimental design from more to less spatial tasks took place in Olton's work as in Gaffan's. The initial experiments were conducted in the radial-arm maze, where the rat's task is to obtain a piece of food from the end of each of the arms. The working memory component of this task comes from the necessity to remember which arms have already been visited within any one trial.

The normal rat shows an impressive capacity to remember which arms it has already visited, even in a 17-arm maze (Olton 1978a). To achieve this high level of performance

it appears not to rely on marking strategies (i.e. odour trails), response chaining, or consistent response strategies (e.g. go to the next arm on the left). When intramaze cues are pitted against the general spatial environment provided by the experimental room, the latter usually dominate performance. There is a recency effect, in that errors increase the closer the first choice of an arm is to the start of the trial. But this recency effect is eliminated if the rat is confined to the centre platform of the maze for 20 seconds after each arm entry. There is no primacy effect in the standard version of the maze, that is, no tendency for the first trial to be relatively immune to later error (but see Kesner and Novak 1982 for primacy and recency effects in a modified form of the test, with hippocampal lesions disrupting primacy, but not recency). If the animal is exposed to several trials within the same day (at a one minute inter-trial interval), there is an increase in errors across trials; but the first visits at the start of each new trial are made with perfect accuracy. Thus, although there is proactive interference between trials, the rat distinguishes between the end of one trial and the beginning of the next. The distribution of errors in space is random.

Based upon these observations, Olton (1978a, p. 365) describes working memory as having 'limited capacity, interference among contents, lack of decay for periods up to several minutes, little if any serial ordering, and no generalization or confusion among items.' If the hippocampus is the site of this memory system, performance on the radial-arm maze should obviously be impaired after hippocampal lesions. But note that the same prediction is made by Gaffan's recognition memory view (since the rat must recognize the locations it has just visited), by O'Keefe and Nadel's spatial memory view (which we discuss in the next section) and several others. Given this convergence of predictions it is not surprising that lesions of the hippocampus, the fornix-fimbria, the entorhinal cortex, or the septal area all profoundly impair performance in the radial arm maze (Jarrard 1978; Olton *et al.* 1978b; Olton and Werz 1978; Walker and Olton 1979). When unilateral destruction of the fornix is combined with unilateral destruction of the entorhinal cortex, performance is essentially normal if the lesions are made on the same side of the brain, but as impaired as with a bilateral lesion if they are made on opposite sides of the brain (Olton 1978b). Thus, for successful performance in the radial-arm maze, the hippocampus has to have access to both its cortical and subcortical connections on at least one side of the brain.

As is the case with Gaffan's experiments, it is possible to eliminate at least one explanation of these results—the one offered by the spatial theory. This was demonstrated by designing a version of the task which depended for its solution on the use of intramaze, non-spatial cues (Olton *et al.* 1979a).

> Intramaze cues were made salient by having the sides of the arms 10 cm high and placing a unique set of visual and somatosensory stimuli in each arm. For example, one arm had black and white stripes in the floor and sides, and a groove 2 cm deep cut into the arm at each black stripe. Another arm had a diamond pattern painted on the sides and a smooth floor and sides. Extramaze cues were minimised by lighting the inside of the maze, covering the top of the maze with cheese cloth and reducing room illumination to a low level. . . . An opaque guillotine door was placed between the centre compartment and the entrance to each of the arms. By lowering the doors, the experiment could confine the rat to the centre.

All rats were first shaped to run out the arms with the maze in stationary position in the room. They were then placed in an 'interchange' procedure as illustrated in Fig. [8.3]. After the rat had chosen an arm and returned to the centre, the guillotine door in front of each arm was lowered confining him to the centre compartment. While the rat was confined, the arms were moved and their topological relationship changed. . . . After the arms were moved, the guillotine doors were raised and the rat allowed to make a choice. When he returned to the centre, the guillotine doors were lowered and the arms were interchanged again. (Olton *et al.* 1979a [our figure reference].)

These procedures ought to have effectively prevented the construction of any kind of spatial map. However, the impairment produced by destruction of the fornix-fimbria is as profound as it is in the usual radial-arm maze (Olton *et al.* 1979a).

Of course, the deficits described so far could be the result not of a failure of working memory, but of a failure to remember the rule on which the working memory task is based (i.e. the reference component of the task). To show a specific working memory deficit, Olton and Pappas (1979) made use of a 17-arm radial maze in which eight of the arms (the same on every trial) were never baited. Thus the rat had two things to remember: (i) the eight arms that were never, and the nine arms that were always, baited at the start of the trial (reference memory); and (ii) the arms in the baited set that it had already visited on a particular trial (working memory). On Olton's theory, damage to the hippocampus should disrupt the latter but not the former. Lesions of the fornix-fimbria severely disrupted the working memory component of this task, as predicted. The reference memory component was also disrupted, but only transiently. Thus an animal with fornix lesions can learn not to enter arms that are never rewarded (Olton and Pappas 1979) and it can also remember which objects have been followed by reward or non-reward (Gaffan 1974).

There are at least two problems with a working memory view of hippocampal function. The first is the severe deficit obtained when rats are required to swim to a submerged platform in a fixed location in the water maze, a reference memory task. An ad hoc solution to this problem would be to argue that determination of one's current position in space requires working memory and, without this, the reference memory task can never be solved. This would appear to be stretching the original idea of working memory

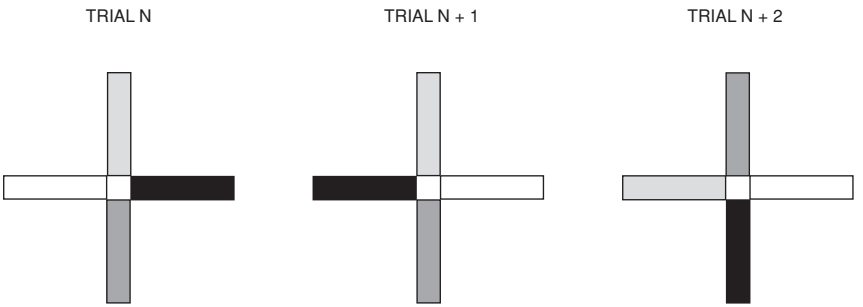


Fig. 8.3 A diagram of the experimental design used in Olton *et al.*'s (1979a) 'interchange' procedure. Each arm of the maze is indicated by a different pattern. After each choice the rat was confined to the centre compartment and the arms were moved while it was so confined.

rather a long way. The second problem is the fact that not all working memory tasks are sensitive to hippocampal lesions. Delayed matching to sample of objects is sensitive to hippocampal lesions only if a limited set of objects is used, and not if trial-unique objects are used. Even trial-uniqueness *per se* does not appear to be the critical factor behind this difference. Rawlins *et al.* (1993) found no deficit with delayed matching-to-sample using stimuli which were not trial-unique but which repeated only seldom, a modest deficit with a single trial-repeating pair of complex stimuli, and a much greater deficit with a single trial-repeating pair of simple stimuli. Following on from this, Yee and Rawlins (1994) found no deficit in delayed non-matching with four choice alternatives when (a) four complex stimuli were used, and (b) four complex stimuli were used with the three to-be-rewarded choices constituting novel stimuli on every trial. But they demonstrated a deficit when (c) four simple stimuli were used. (Comparison of (a) with (b) demonstrated the presence of interference effects which did not appear to be exacerbated by the lesions; we discuss this point further below.)

As with the other theories we have considered, this is a limited domain account (e.g. it cannot account for increased resistance to extinction after hippocampal damage) which does not incorporate any attempt to map the details of the theory to the known anatomy and physiology.

8.12 TEMPORAL BUFFER

A theory superficially similar to the working memory theory is that the hippocampus constitutes a high-capacity, intermediate-term memory store (Rawlins 1985). Rawlins' approach has a major similarity to our own, in that he argues strongly against 'the idea that the hippocampus is required to process items only of some qualitatively specifiable kind, and is not required to process items of some complementary set. In contrast, it is now proposed that the hippocampus is needed to process stimuli of all kinds, but only when there is a need to associate those stimuli with other events that are temporally discontiguous' (Rawlins 1985, p. 479).

The basis of Rawlins' theory is a careful analysis of the tasks which have purported to distinguish different types of memory, with the conclusion that 'the sensitivity of these tasks as indicators of hippocampal dysfunction is not well predicted by existing task classifications based on current views of hippocampal function. . . . Existing theories need to be significantly modified. They need to explain why increasing or decreasing the degree of temporal discontiguity within their paradigmatic designs concomitantly increases or decreases the tasks' sensitivity to hippocampal dysfunction.' (Rawlins 1985, pp. 496–7). With these conclusions we concur.

Rawlins goes on from this to the stronger claim 'that the hippocampus functions as an intermediate memory system. Information is normally admitted to this store in parallel with admission to a much more limited short-term store. When the intermediate-term system is incapacitated, any task in which information must be given temporary storage will become insoluble once the storage duration or the number of items to be stored exceeds the residual short-term storage capacity. This view holds for information of all kinds and suggests that the critical determinants of task sensitivity to hippocampal

dysfunction are temporal discontiguity and number of items requiring temporary storage.' (Rawlins 1985, pp. 496–7).

We remarked in a commentary on Rawlins' paper that his 'reanalysis of previous experiments and the results of his own experiments make an impressive case for the importance of temporal discontiguity in the effects of hippocampal lesions. However, his success in substituting temporal discontiguity for previous constructs prompts the question: does temporal discontiguity itself mask some other factor? Rawlins himself identifies one possibility. He states that "the time over which storage is required *and the number of items needing storage* can be seen as trading off against each other" (our emphasis). However, increasing temporal discontiguity increases the probability of interference from extraneous stimuli' (McNaughton 1985, p. 508). McNaughton (1985) then reviewed data on passive avoidance, extinction, and taste aversion conditioning, all of which were incompatible with Rawlins' theory. Since then, Rawlins' own laboratory has produced data which effectively demolish a time–capacity view of the hippocampus: hippocampal lesions impair delayed matching-to-sample most with a pair of simple stimuli, less with a pair of complex stimuli, and not at all with pairs of pseudo-trial unique stimuli, distinctions that do not involve the temporal domain (Rawlins *et al*. 1993).

Unlike many of the other theories we have looked at, Rawlins' theory is not intended to be limited in its domain, although it was not supplied with any detailed physiological or anatomical machinery. However, from the data briefly reviewed above, we must conclude that, while the strong parametric influence of time on the effects of hippocampectomy must be accounted for by any successful theory, time itself is not of the essence. We return to this point below.

Rawlins' view of a temporally limited intermediate store is very similar to an idea that allows many of the 'memory type' (hippocampal vs. non-hippocampal) theories to account for the limited retrograde effects seen in amnesics. For these theories, the 'hippocampal store' is not strictly limited in the time that it can hold information. Rather, the hippocampus is specialized to hold certain types of complex information, and does so for as long as is necessary for cortical circuits to consolidate that information. The great variety of proposed memory typologies of this kind, noted above, suggests that they may all suffer from some kind of fundamental intractability. But this is not the only reason to doubt the validity of this approach—serious problems are raised by the nature of the processes which must be attributed by this class of theories to the 'hippocampal memory store'.

H.M. has an intact long-term memory for information acquired pre-operatively. It follows that the putative hippocampal-dependent memories are distinct from non-hippocampal memories not only in terms of their 'type' but also in that if they require storage in the hippocampus, they do so only for a relatively brief period of time. Current memory theories of the hippocampus therefore require (usually implicitly) that hippocampal memories are ultimately copied into cortex (or wherever is their final repository). Wilson and B. McNaughton (1994), for example, explicitly in their case, suggest that the hippocampus contains a fast flexible store which can then replay information into cortical circuits during consolidation (particularly during sleep; see Appendix 7). This idea of a specialized store which ultimately transfers its information to other parts of the brain can certainly account for an initial sensitivity of only certain classes of

memory to hippocampal damage, and for the later lack of such sensitivity. It is also a superficially attractive notion when one thinks of the ease with which areas of memory can be copied in computers.

However, the idea becomes less attractive if we ask *why* computers need to copy memory information, and whether what we know of neural plasticity is in any case compatible with such copying. Computers copy 'memories' from one part of the machine to another with great facility. Furthermore, hippocampal and cortical synapses both show long-term potentiation, and this could, in principle, store information in a computer-like fashion. But the analogy breaks down when we note that copying of memory areas, one to the other, is easy in a computer because, in memorial terms, it is unnecessary. Memory is copied when one area is to be modified and another is required to retain the original information. But the process is only possible because *both* areas are equally good repositories of the original information (which is coded by bit pattern rather than location); and what is copied to where is completely arbitrary. The analogy with the brain breaks down in two separate ways. First, a supposed significant difference in the nature of hippocampal and non-hippocampal storage is the only reason for postulating a particular 'type' of hippocampal memory in the first place. *Ex hypothesi*, therefore, the uniformity between memory areas in the computer is not true of the brain. Second, the brain codes information positionally (the point in the brain, for example visual cortex, defines the nature of the information), not in terms of bit patterns.

Let us consider the formation of simple, hippocampal-independent, memories. There is good evidence (Chapter 6; Appendix 2) that active avoidance depends on long-term potentiation (or some similar process) of input from sensory systems to the amygdala. We can also (Appendix 6) note that conditioning of the nictitating membrane response involves plasticity of sensory inputs to cerebellar circuits. Much of what we term memory must, therefore, involve CS–response program associations, produced by long-term potentiation or something like it, in a bewildering variety of different places in the brain. Because hippocampal-dependent memories are not sensitive to hippocampal lesions once they are fully consolidated, it follows that even these types of memory must ultimately be instantiated in strengthened connections between and within appropriate extra-hippocampal areas. We can ask, then: if the cortex (or other final repository of long-term 'hippocampal' traces) can store the relevant information after consolidation, why is the hippocampus needed in the first place?

There is a further fundamental logical flaw in the idea of the hippocampal formation as an intermediate memory store which follows from the general nature of hippocampal-dependent memories. Whatever the specific type postulated to be hippocampal-dependent, they all have in common that they are seen as being more complex than simple associative memories, and usually (most clearly in the case of relational, declarative, contextual, and configural theories) that they involve combinations of elements which can enter separately into hippocampal-independent associations. This leads, given the nature of non-hippocampal storage of information, to the conclusion that the hippocampus should be bigger than the totality of those parts of the brain with which it interacts in the formation of memories. Not only is this clearly not the case, but the relative increase in size of the hippocampus with encephalization has been nothing like as great as the increase in the size of the cortex.

Let us consider this problem from the point of view of relational memory (as one of the most recent, coherent, and successful memory theories of hippocampal function). Cortex (or another non-hippocampal area) is held to support simple associations of the 'B will follow A' type, which we can represent as A–B. It is only when some specific, more complex relation (e.g. A–B–C, as distinct from A–C–B, B–A–C, B–C–A, C–A–B, and C–B–A) must be encoded that the hippocampus is required. Note that if the hippocampus is to be the location where such relational information is encoded, this implies that the hippocampus can duplicate all the possible associative connections of those other parts of the brain that deal with the same information but at a non-relational level (since A–B and B–C are elements of A–B–C). This type of theory also implies that the hippocampal input and output connections can, in principle, carry the factorial combination of all the embedded simple associative connections taken N at a time, where N is the largest number of simple elements normally combining to form a distinct relational memory. This is inherently unlikely. With minor changes of detail this same argument can be applied to all the current 'memory type' theories.

We are not arguing against the idea that there are iconic, short, medium, intermediate, and long-term stages of information storage on the way to production of the long-term trace. At the biochemical level these stages are clearly present (Rosenzweig *et al.* 1993). However, they reflect stages of electrical or biochemical cascades at a single location (node) rather than the transfer of a trace from one location to another. Nor are we arguing against the idea that information can be transferred between brain areas and form new durable traces. This must occur when memories are integrated to form novel concepts (but here the new information is of a form quite different from the old). It must also occur with interhemispheric transfer (but here the areas between which the transfer is taking place are essentially isomorphic, and the transfer is required because of lateralization of inputs or outputs, not because either area is, in principle, incapable of forming the associations initially encoded in the other).

We are also not arguing against the distinction between working and long-term memory. Working memory can be conceptualized as depending on an interaction between the frontal cortex and other areas of the brain (Chapter 6; Appendix 3). It reflects recursive processing (rather than a process like long-term potentiation) which can be viewed as involving a 'working' component supported by activity in the frontal cortex and a 'memory' component, resident within the area of cortex currently interacting with frontal cortex. Thus, the specific details of the information being processed depend on the same neural circuitry as was activated by the original perceptual information now actively being recalled. Here too, then, we can view memory itself as unitary, with working memory reflecting the reactivation by frontal cortex of the circuitry which contains the long-term trace (in a distributed form). Critically for hippocampal analysis, working memory is clearly associated with frontal cortical function (see, for example, Funahashi and Kubota 1994; Selemon *et al.* 1995) and (except incidentally, as we discuss below) is not associated with hippocampal function.

The role of the hippocampus, then, is not in our view to store information temporarily and then transfer it to cortical areas during consolidation. Rather, it interacts with storage, which is always cortical, to allow consolidation of the cortical traces to proceed without degradation.

8.13 SPATIAL MEMORY

The earliest theory of hippocampal function which attempted to integrate the bulk of the anatomy, physiology, and psychology of the hippocampus was O'Keefe and Nadel's (1978) spatial theory of hippocampal function.

This was, and remains, an impressive piece of work. Its evidential base is much greater than that of most of the other theories covered in the present chapter: single-unit recordings showing place fields of hippocampal cells (reviewed in Chapter 7); the extreme sensitivity of spatial tasks to hippocampal lesions; a coherent account of the role of theta in hippocampal function; consideration of the detailed anatomy of the hippocampal formation; and derivation of the underlying concepts from the philosophical works of Newton, Leibniz, Berkeley, and Kant. It has also achieved greater longevity than many of the other theories discussed here. The full version of the theory (which had been developing for some time previously) was published in 1978 and it is still being strongly defended now, over 20 years later.

O'Keefe and Nadel's theory is based on

> the view that the hippocampal system is quite specifically a spatial mapping system. [But] the cognitive map theory of hippocampal function is, first and foremost, a theory about *memory* for spatial layouts and the ways in which animals use such a memory system for adaptive behaviour in the world. . . . Adherence to such a multiple memory system view has now become the accepted norm [but at that time our view] was sufficiently different from the standard view that cognitive map theory was considered merely a theory about space, not memory.
>
> We assumed that there were two classes of learning systems, corresponding roughly to the notions of 'knowing that' and 'knowing how'. We created names for there two classes of learning systems: *locale* and *taxon*. These names were chosen to represent certain critical features of each of the learning systems. In the case of the locale system this was the fact that the memories it contained incorporated information about place and time; in the case of the taxon systems it was the fact that they were governed by the rule of category inclusion. Memories in taxon systems were organized in terms of features and their similarity/dissimilarity, while those in the locale system were not. There was also an important distinction between these learning systems in terms of the role of time/place context—it played a central role in locale-based memories, but was absent from taxon-based memories.
>
> The foundation for . . . the locale system lies in its inclusion of spatial information. In the years since our formulation of this theory, there has been much support for the view that the hippocampus is crucially important in the processing of spatial information. What has been rejected is our insistence that this spatial signature is the best way of describing hippocampal function.
>
> Finally, and very briefly, I will toss in the towel and admit, as we did in 1978, that at least in the case of the human hippocampal system, there is more than merely spatial mapping going on. . . . Why not admit [also] a more abstract representational system into the monkey, rat, or even bird hippocampus? The answer, then as now, is simply that we did not, and still do not, see the need for such hypothesis drift. (Nadel 1991, pp. 221–2.)

The range of the home database for the spatial theory is impressive. At the time of its proposal this range was so much greater than that of any other theory that a few problematic items around the edges (e.g. accounting for the effects of hippocampal lesions on DRL performance) did not detract from its appeal. There are also problems with the

detailed machinery required by the theory; but, even now, this should not count too strongly against it, since few of the other theories suppose any such detailed machinery at all.

Now, however, there are competing theories (e.g. Eichenbaum 1996) which have been developed in response to perceived deficiencies of O'Keefe and Nadel's theory but which, at the same time, attempt to incorporate the spatial results. The purely spatial version of O'Keefe and Nadel's theory can be viewed as a special case of the relational theory. It might even be argued that in its 'cognitive map' as opposed to purely spatial form, O'Keefe and Nadel's theory is simply a variant of the relational theory. However, there is one important point of difference. According to O'Keefe and Nadel, the hippocampus is itself the place where the map is stored. It is difficult to reconcile this with the limited retrograde amnesia observed in subjects such as H.M. This would surely require, in addition to the hippocampal map, considerable extra machinery for transferring maps from hippocampus to neocortex while also assuming that the cortex has a similar mapping capacity, in which case the hippocampus, as hypothesized, should not be strictly necessary (see the discussion in the previous section).

More importantly, animals with hippocampal system damage can acquire some forms of spatial task and fail to acquire some forms of non-spatial task. Within the memory domain the clearest non-spatial tasks affected by hippocampal system damage are various forms of delayed matching and delayed non-matching in which, except in the special case of matching or non-matching to position, spatial strategies are deliberately excluded as a means of solution.

Intact spatial learning is shown in hippocampal lesioned rats under interesting circumstances. Eichenbaum *et al.* (1990) trained fornix-lesioned rats in the standard water-maze apparatus, but starting them always from the same position rather than from randomized positions. The lesioned rats learned the task as well and as quickly as controls. When switched to the conventional randomized start procedure, the controls immediately swam directly to the platform, but the lesioned rats demonstrated a deficit (although this was not as large as the deficit observed with similar lesions when acquisition is with randomized multiple start positions). A crucial point (not emphasized by Cohen and Eichenbaum, 1993, when discussing these data) is that 'performance of fornix rats was indistinguishable from that of sham rats on the standard transfer test; both groups spent most of their time in the [correct] half of the pool . . . [with] no significant difference between groups in the distribution of swim times among quadrants' (Eichenbaum *et al.* 1990, p. 3535). From this we can conclude that the fornix-lesioned rats had learned the position of the platform in a truly spatial fashion, as opposed to simply learning to swim towards some specific cue (taxon learning in O'Keefe and Nadel's terms). But when they were required to approach the platform from novel directions they suffered significant problems. It could be argued that it is the representational flexibility required by multiple starts that causes the problem. However, Whishaw (personal communication) has shown that fornix-lesioned rats can learn even the multiple start version of the water maze as well as controls provided that their initial training prevents them from making errors (and hence presumably adopting incorrect initial response strategies). In this experiment, the rats were initially placed on the platform, then placed in the water close to and facing the platform, and so on, with farther starting positions on each trial, until they had become the conventional starting points at the edge of the pool.

We may appear to have dismissed O'Keefe and Nadel's theory in a somewhat cursory fashion. However, the reasons we have found (see also Appendix 8) to reject a *purely* spatial view of the hippocampal formation are ones which have also compelled Nadel's own, essentially similar, view quoted above, except that he does not extend the view to non-human animals. Whether we should also reject their fuller cognitive map theory is another matter. It is likely that fairly modest updating of the theory could bring it into line with the modern data. The issue of whether such an updated 'cognitive map' theory would differ from our theory except in terminology will have to await its appearance. Certainly, the capacity to distinguish between a generalized cognitive map and the behavioural inhibition approach is limited. A relatively high resistance to change of taxon-based as opposed to locale-based hypotheses (the animal's, that is, not the experimenter's) is a crucial axiom of O'Keefe and Nadel's theory (see, for example, O'Keefe and Nadel 1978, p. 338 et seq.) and is consistent with some of the detailed effects of septo-hippocampal damage (Janis *et al.* 1994). However, once space itself is removed as the defining characteristic of locale hypotheses, it is not clear how we distinguish the predictions made by the O'Keefe and Nadel theory from those that emerge from the hypothesis of a simple difficulty in changing previously established response tendencies.

Whatever the merits of a uniquely spatial view of the hippocampus, any alternative theory of hippocampal function must provide an account of the overwhelming impression from many studies that the hippocampus is concerned with space, and also for the data indicating that hippocampal neurons are capable of forming place fields (which we reviewed, and interpreted somewhat differently from O'Keefe and Nadel, in the previous chapter). It should also be emphasized that O'Keefe and Nadel's is one of the few theories in the area that has been designed as a modern theory should be: it attempts to account for all the data, not artificially restricting the explicanda to those few which can be accommodated by the theory; it provides detailed anatomical and physiological machinery, and a prescription for the functions of each part of the machine; it allots roles to long-term potentiation and the theta rhythm (if anything it puts too much functional weight on theta, treating this as an essential index of the spatial map). In these ways it provides a means to falsify the details as well as the main thrust of the theory. This contrasts starkly with many current theories which simply ignore the theta rhythm and other aspects of the relevant anatomy and physiology altogether and so cannot be falsified by data from these sources. What seems so surprising is that, with O'Keefe and Nadel's theory as a glowing example, the vast majority of more recent theories have gone back to the 'bad old days' when a theory was often stated in a single sentence, applied to only a carefully selected set of data, with no specification of the detailed underlying machinery at all.

8.14 INTERFERENCE: SOME INITIAL COMMENTS

The final issue we consider is that of interference, an issue that has been present implicitly or explicitly in many of the theoretical approaches discussed above. Rawlins' (1985) amalgam of time and memory capacity implies a link with interference. Hirsh (1974), O'Keefe and Nadel (1978), Olton (1978a), Winocur (1981), and Gaffan (1994) all

explicitly include interference as a main factor leading to hippocampal sensitivity of a task. We consider shortly Warrington and Weiskrantz's theory of the hippocampus as functioning solely to deal with interference. In our view, they abandoned this theory too soon.

The topic of interference must be approached with great care for two quite separate reasons.

(a) The absolute amount of interference present in a task is usually not measured and also depends on how interference is defined. The temptation, then, is to wave the magic wand of 'interference' post hoc in a way which could explain almost any result but can predict none. This problem can most easily be circumvented by taking a paradigm which is insensitive to hippocampal lesions and creating task variants which all remain exemplars of the paradigm but which, nonetheless, have increasing amounts of interference. Somewhat less easy is to identify sources of interference in tasks of a hippocampal-sensitive paradigm and remove them (as has been achieved with the hippocampal-insensitive versions of the Morris water maze). However, few paradigms are consistently affected by hippocampal lesions (Gray and McNaughton 1983), giving one a starting point for comparisons which could determine, in each case, the underlying source of hippocampal sensitivity. For both nominally sensitive and nominally insensitive paradigms, then, the role of interference can be separated from the role of the process officially tested by the paradigm.

(b) Interference can either be part of the description of a body of results—requiring an explanation—or it may have the status of a theoretical construct, constituting the explanation. Thus, we shall consider shortly what is, in essence, an interference theory of hippocampal function. However, interference theories of the hippocampus (if they are to account for its selectivity) must attribute to it the *removal* rather than the generation of interference. They must, in consequence, ascribe some specific, potentially quite detailed, processing function to the hippocampus from which one may deduce a specific deficit in amnesic performance by saying *both* that the task contains interference *and* that the effects of this interference are normally eliminated by an intact hippocampus.

Weiskrantz and Warrington (1975) proposed a hypothesis to account for the detailed aspects of the amnesic syndrome which is very close to being a pure interference view. According to this hypothesis, amnesics do not forget or fail to consolidate what they learn (a view which is much better substantiated now than it was then). Rather, they cannot select the correct response out of a range of competing possibilities. Thus Warrington and Weiskrantz (1968), testing memory for repeated lists of words, found that amnesics generated more false positives than controls and that 50 per cent of these errors were intrusions from previous lists. Also, from an interference point of view, the beneficial effects of cueing for recall are due to the limits the cue places on the range of possible false positives (Warrington and Weiskrantz 1974). When the possible alternatives were reduced to one (by prompting with a set of three letters which can be completed by only one word in the English language), the amnesics actually remembered better than controls over a retention interval of 24 hours (Warrington and Weiskrantz 1978).

In apparent contradiction of this hypothesis, however, as shown in Fig. 8.4, Warrington and Weiskrantz (1978) found that amnesics were not impaired relative to controls on the

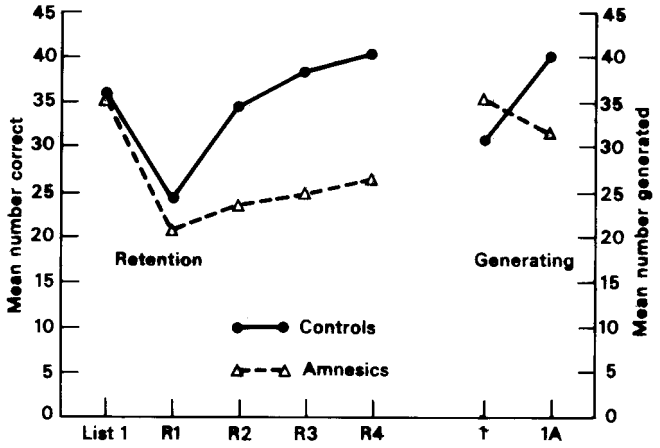


Fig. 8.4 Left panel: reversal learning. For each recall cue (the initial three letters of a word) there were only two English words available as possible responses. Subjects were first taught one set of words (List 1) and then were given four trials with the alternative set (R1, 2, 3, 4). The same cues were used on all five trials. Right panel: generating scores. At the end of the reversal learning stage subjects were asked to produce two words in response to each cue. 1, generated words which came from list 1; 1A, generated words which came from the reversal list. (From Warrington and Weiskrantz 1978.)

first reversal of four successive reversals, although they were impaired subsequently. This appears problematic for a simple interference theory (and, indeed, any simple behavioural inhibition theory, see below), since an impairment based on excessive interference from items learned in the first list would have been expected to be greatest on the first reversal trial. This result therefore drove Warrington and Weiskrantz (1978, p.174) to suggest that 'it is possible that prior responses, while not stronger initially, may extinguish more slowly when a new response is required to the earlier cue, or indeed may decay more slowly over longer intervals even when no response is demanded' in amnesic subjects.

Such an ad hoc explanation of the data is required, however, only if we view amnesic subjects as somehow receiving more interference than controls. But, if we view the hippocampus as a device which, by whatever means, can negate the interference experienced by both controls and lesioned subjects, then Warrington and Weiskrantz's (1978) results do not create the same problem. Inspection of Fig. 8.4 shows that on trial R1, both the controls and the amnesics are equally impaired—presumably because of interference. However, fairly rapidly over trials R2,3,4 the controls become capable of suppressing this interference, whereas the amnesics develop this suppression to a much lesser extent. Thus, and this will be important for our theory, amnesics are not more susceptible to interference when this occurs, but rather are less capable of subsequently suppressing the prepotent, incorrect competing responses generated by such interference. As a result, a deficit will only be observed in those cases where controls are successful in suppressing interference to some extent, but not in those cases where both patients with hippocampal damage and controls are affected by interference, the effects of which the controls are unable to suppress. Consistent with this view, Baddeley and Wilson

(1994) showed that, with 'errorfull' and 'errorless' training procedures, amnesics were not so much insensitive to the presence of errors as incapable of learning to eliminate them (this was particularly evident across rest breaks, where there was even some suggestion that amnesics' performance could deteriorate with training).

There are specific demonstrations that rats with hippocampal lesions, like human amnesics, are more susceptible to interference. Thus Jarrard (1975; see also Thomas 1978, for the effects of fornix-fimbria lesions) showed that hippocampal lesioned rats were impaired in a delayed spatial alternation task if they were forced to run in a running wheel during the delay interval, but not otherwise. Such animals are also more susceptible to interference occurring between acquisition and a retention test. Winocur (1979) trained rats on a simultaneous discrimination between horizontal and vertical stripes. During the 5-day training–retention interval, animals were given one of three kinds of experience: an insoluble problem with diagonal stripes and random 50 per cent partial reinforcement; reward for all choices with no stimuli present; or rest. Hippocampal-lesioned animals were impaired in retention of the original habit only after the intervening insoluble problem. Note, however, that pre-training on the insoluble problem also gave rise to an impairment in the lesioned animals in initial acquisition of the horizontal–vertical discrimination; thus Winocur's (1979) findings indicate that hippocampal lesions cause a general susceptibility to interference, not a retention deficit as such. Similarly, Winocur (1985) found that hippocampal lesions increased the interfering effects of the intrusion of a reinforced lever into a go–no-go task at both 0 s and 10 s delays. The importance of interference is also supported by the lack of effect of hippocampal lesions on delayed matching to sample with trial-unique objects. In this paradigm, the object's uniqueness eliminates the interference which could occur with trial-repetitive objects (see section on working memory above).

As with the human case (Warrington and Weiskrantz 1978, discussed above), there is evidence in animal experiments that interference effects from competing response tendencies can build up. Meck *et al.* (1987) investigated responding on a peak interval procedure, in which rats show responding locked to a particular time at which reward occurs and in which hippocampal damage shifts responding to earlier times. When this task was learned pre-operatively, the full effects of fornix-fimbria lesions developed only over a number of sessions. Meck (1988) showed that the controls, but not lesioned animals, in this task show 'hunting' of their responses. This indicates a tendency to switch between competing response tendencies on different trials. In another experiment able to bear a similar interpretation, Yee and Rawlins (1994) compared complex and simple stimuli in a delayed non-matching to sample task. Animals with lesions of the hippocampus or fornix were impaired only when the simple stimuli were used. The introduction of the simple discriminanda produced an equal impairment in performance of both the control and lesioned rats, but the controls improved their performance over trials whereas the lesioned rats did not.

Yee and Rawlins (1994) also explicitly tested, in a separate experiment, for the effects of interference *per se*. They used the infrequently repeating, complex stimuli with which they had shown unimpaired delayed non-matching to sample. With these, they investigated the effects of replacement of the target stimulus shown on the information trial with completely novel stimuli on the choice trial. This manipulation significantly

improved performance, indicating an effect of interference in the non-replacement condition, with a similar improvement in control and lesioned animals. This result needs to be treated with some caution, as the control data could well have been restricted by a ceiling effect and there was no difference between the two behavioural conditions in the hippocampal animals at the end of training. Given the slow development of the lesion effects considered in the paragraphs above, this type of experiment clearly needs replication, but, for the moment, the bulk of the data appears to be consistent with the view that hippocampal lesions do not exacerbate the effects of interference *per se*, but rather, when control animals have some means of eliminating those effects or compensating for them, hippocampal lesions impair this suppression of errors.

So far, we have looked at an exemplar of interference theories and some explicit tests of the effects of introducing interference on sensitivity to hippocampal lesions. In the following sections, we investigate the explicit and implicit role of interference in relation to each of a number of types of 'hippocampal' memory, as well as tasks which have only minimal memory components.

8.14.1 **Interference and context**

Winocur's (1981) hypothesis that non-human subjects with hippocampal lesions are excessively dependent on contextual cues has also been applied to the human amnesic syndrome (Kinsbourne and Wood 1975; Winocur 1981). It is difficult, however, to distinguish this approach from Weiskrantz and Warrington's (1975) interference hypothesis.

Consider, for example, an experiment by Winocur and Weiskrantz (1976) in which amnesic subjects were tested on lists of paired associates (which they are usually very bad at learning), but with the possible set of responses constrained by a rule common to the whole list (e.g. that all pairs rhyme). One can regard the rule as providing partial information (which reduces interference on the model of the cued recall experiments described earlier); or, equally naturally, one can regard the rule as providing a context, namely the one made up of the set of words with the same sound as the stimulus word. On either view, amnesics should learn a list made up in this way more effectively than usual, and this in fact was the case (Winocur and Weiskrantz 1976). As in Warrington and Weiskrantz's (1974, 1978) experiments on reversal learning, the patients were severely impaired when required to learn a second list sharing the same rule and initial words as the first, and they showed a strong tendency to produce false positive responses from among the associates learned in the first list.

A second experiment (Winocur and Kinsbourne 1978) derived more explicitly from the concept of context. Amnesics were tested on paired associate learning under one of two conditions: in a standard experimental room with normal sound and illumination levels, or in a room illuminated only by a bright red lamp and with background music. When both learning and retention were tested in the more unusual environment, the amnesic deficit was greatly reduced. Conversely, if amnesics learned two successive lists in the two different rooms, this reduced the degree to which learning of the first list impaired subsequent learning of the second list. In neither experiment was the performance of normal subjects affected by the context of learning. These results parallel those reported by Winocur and Olds (1978) in hippocampal-lesioned rats: the amnesic, like

the hippocampal, deficit is reduced by a strong context common to the conditions of learning and recall; the reversal deficit is reduced by a change in context between learning and reversal.

In sum, the provision of a strong cue as to the correct response aids the amnesic to select it from among the possible competitors. But there is no good reason, either empirical or theoretical, to distinguish between the cases in which the cue is intrinsic to the item to be retrieved (cued recall by partial information) or extrinsic to it (contextual cueing). There is also no good reason to see the cueing process as being memorial. For example, Warrington and Weiskrantz (1970, 1974) found that if cues were provided only at the time of learning, recall was not aided; but if they were provided only at the time of retention testing, the amnesic deficit was greatly reduced. Thus, cued recall aids retrieval, implying that the material has been stored, though not necessarily stored normally.

On this basis, contextual cues (i.e. contextual in Winocur's as opposed to Hirsh's sense) can be viewed as important because they reduce the effects of interference; but they are not the only such cues that can be so used, and interference appears, therefore, to be the more general concept.

8.14.2 **Interference, recognition memory, and working memory**

The experiments on non-human animals by Olton's group and by Gaffan, described earlier, seemed to bring us within hailing distance of the amnesic syndrome described after hippocampal damage in human beings. However, this is only if the interpretation once placed on that syndrome, that it is due to a complete and specific loss in the ability to form long-term memories (Milner 1968), is wrong. Fortunately, this interpretation cannot apply to working and recognition memory deficits.

Careful analysis of the working and recognition memory tasks suggests that hippocampal deficits here are not the result of a direct failure of memory at all. The key factor which brings out a memory deficit in these types of task seems to be the existence of potent sources of interference. This is clearly apparent in Gaffan's (1977b) experiments on delayed matching-to-sample of colours and spatial positions. The deficit shown by the fornix-lesioned monkeys was confined to the item presented first in the list, the one most likely to suffer from retroactive interference (see also Kesner and Novak 1982).

There are other features of the data which argue for interference and against Gaffan's hypothesis that the hippocampus subserves the discrimination of familiarity. The first is the observation that amnesics are particularly strongly wedded to the first response that they learn. If items have been presented with equal frequency the first item should in many cases be the least familiar. Second are the results of a supplementary test in Warrington and Weiskrantz's experiment on reversal learning (see above). After reversal learning was complete, the subjects were asked to produce two words in response to each cue (recall that the cue consisted of the first three letters common to only two words in the English language). As shown in the right-hand panel of Fig. 8.4, amnesics produced significantly fewer list 2 responses, although they actually produced more list 1 responses than controls. But if the amnesic deficit lies only in recognition, as proposed by Gaffan, or only in working memory, as proposed by Olton, there should be no difference between amnesics and controls in generating items from either list.

The use of trial-unique as opposed to trial-repetitive objects in a working memory task such as delayed matching-to-sample should reduce interference, and, as predicted, this procedure eliminates the lesion deficit in the conventional form of this task. Presumably for the same reason, hippocampal-lesioned rats may show an impairment in delayed non-matching-to-sample if the information (lever-press) response is rewarded, rendering it more similar to the choice trial, but not if it is not rewarded (Rawlins and Tsaltas 1983). In this latter experiment introduction of the reward itself reduced both control and lesioned animals to chance responding (presumably as the result of response competition), while the introduction of a fixed ratio (FR) 5 and then FR10 reinforcement schedule for lever-pressing on the information trial (rendering this more distinctive again) improved correct responding in all groups, but differentially improved control responding.

8.14.3 **Interference and spatial memory**

As delineated by Nadel (1991), the spatial memory theory of hippocampal function is intended to be purely spatial only in the case of non-human animals, or possibly non-primates. That is, in the case of human beings the functions of the hippocampus are presumed to be lateralized, with one hemisphere producing 'cognitive' rather than 'spatial' maps. However, even 'in rodents, lesions of hippocampus or related structures impair odor discrimination learning, the ability to time a short interval, and performance in configural discrimination tasks that require remembering unique combinations of stimuli' (Squire and Cave 1991, p. 269). We shall, therefore, consider the role of interference in the more general case of cognitive mapping below when we consider relational memory tasks.

In the specifically spatial case, it is clear that, where a spatial task is normally sensitive to hippocampal lesions (excluding therefore, for example, simple position discrimination in a T-maze), there will be ample opportunity for interference from the many different alternatives available in the task. More importantly, there are a number of cases (see above, Section 8.13) where hippocampal-lesioned animals could show apparently normal spatial location learning when training procedures were used which were intended to remove the effects of competing response tendencies. It could be argued that these procedures have encouraged taxon as opposed to locale learning and so rendered the tasks non-spatial (except that probe tests suggested otherwise). For our present purposes the critical point, however, is that removal of presumed sources of interference rendered the tasks insensitive to the lesions.

8.14.4 **Interference and temporal memory**

As we have discussed already, in many tasks increasing delay coupled with increasing information load will result in increasing interference. In most cases, the separate contributions of time and interference will be impossible to disentangle. However, there is at least one clear case where delay does not usually result in a hippocampal deficit: taste aversion conditioning. Here the delay over which conditioning can be obtained is itself unusual, and there is good evidence that this is because most other paradigms have much

greater levels of inherent interference (Revusky 1971). For this reason, we predicted that 'taste aversion would become sensitive to hippocampal lesions if the animal were presented with a variety of non-novel tastes during the conditioning interval' (McNaughton 1985a, p. 509). To our knowledge this prediction has not yet been tested.

8.14.5 **Interference and relational memory**

We discussed Eichenbaum and Cohen's (1993) concept of relational memory in some detail in Section 8.7. Here we take a somewhat more general view of relational memory, subsuming not only their view but also cognitive mapping and similar theories. The key point from our discussion of their theory was the argument that the relation A–B–C can only be formed if A, B, and C are concurrently activated; and, hence, there is a high probability that relations such as B–A–C and C–A–B will be formed incidentally. As a result, relational memory tasks are likely to be high-interference tasks.

8.14.6 **Interference and behavioural inhibition**

The idea that the hippocampus mediates behavioural inhibition is not strictly a memory theory; nor, as we saw in Chapter 6, can it account for all of the effects of hippocampal lesions. But even a fairly loose idea of behavioural inhibition encapsulates a feature of the data often ignored by memory theorists and which certainly presents unusual difficulties for such theories. As we saw in Chapter 6 (see also Appendix 8), the animal literature has many examples of cases where learning of a simple task is intact in hippocampal lesioned animals but where unlearning, which should be as simple or even simpler if considered purely in terms of the learning requirements themselves, is impaired.

Human amnesics have not usually been tested in the tasks which show apparent losses of behavioural inhibition in animals. However, the animal data themselves show that behavioural inhibition *per se* is not lost after hippocampal lesions. Thus, hippocampal-lesioned animals can, for example, learn a DRL task, provided that no prior training has been given on a competing response. Similarly, fornix-lesioned rats can learn locomotor passive avoidance as quickly as controls, provided the initial baseline tendency to respond is low, whereas they show a deficit when transferring from active to passive avoidance dependent on the same response, i.e. if they have to shift from producing the response to inhibiting it (Okaichi and Okaichi 1994).

Thus, the requirement for behavioural inhibition *per se* is not critical; it seems that an intact hippocampus is required only when substantial interference from a competing alternative must be eliminated.

8.14.6 **Interference as separate from memory**

Pursuing further a theme initiated in the preceding section, if interference is a critical element of tasks sensitive to hippocampal lesions, its effects are not limited to memory tasks. On the contrary, we have seen several examples of a heightened susceptibility to interference in lesioned animals under conditions in which memory plays a minimal role (Chapter 6; Appendix 8). Indeed, in some experiments the difficulty experienced by the

lesioned animal seems to arise because it is *more* strongly dependent than intact controls on what it has already learned. Thus, if anything, these experiments suggest that the lesioned animal forgets too little rather than remembers too little.

This is obvious in the case of extinction and reversal deficits. However, a more subtle interference was demonstrated by Winocur and Mills (1970). They showed that hippocampal-lesioned rats were impaired in learning a simultaneous pattern discrimination only if they had first learned a brightness discrimination. Another example is Donovick *et al.*'s (1979) experiment on the effects of septal lesions. Rats were first trained on a simultaneous discrimination with either brightness or position as the relevant dimension. The other dimension was then used to add a second, redundant cue. Transfer of behavioural control to the added cue was impaired by septal lesions. A third example is Kimble's (1975) analysis of the behaviour of hippocampal rats under various conditions of reinforcement in a Y-maze with one arm lit and the other unlit. Kimble scored his subjects as showing position or brightness hypotheses, depending on the consistencies linking their successive choices of arm. Hippocampal-lesioned rats clung longer to whichever response strategy they first happened upon, and their subsequent behaviour in a changed environment was dominated by this strategy to an abnormal degree.

The necessity to determine how the control animal is solving a problem is brought into sharp relief by Eichenbaum's experiments in which fornix lesions affected simultaneous discriminations with olfactory stimuli, whereas such discriminations are usually unaffected by hippocampal damage when other stimulus modalities are employed. With simultaneous olfactory discrimination there was 'a qualitative distinction between normal and [fornix-lesioned] animals in their response latencies. Normal animals had bimodal distributions of latencies, with one peak corresponding to left-port responding and the other peak to right-port responding, suggesting that the animals sample both ports and make some comparative judgement at the time of their response decision. By contrast the responses of animals with hippocampal system damage had shorter latencies and a single-peaked distribution, suggesting that these animals made their choices without comparative judgements at the time of their response decision' (Cohen and Eichenbaum 1993). Thus, in 'simultaneous' olfactory discrimination there is in practice a strong successive or conditional element in control rats' responding, and the evidence shows that the lesioned rats fail to inhibit responding, fail to correctly sample and assess the alternatives, and hence make errors. (It is also possible, however, that olfactory stimuli are processed differently than other modalities and, in particular, more easily elicit affective associations; Herz and Cupchik 1995.) Brown (1992) made a similar analysis of the detailed behaviour of rats in the radial-arm maze. He found that control animals oriented towards arms and made a number of 'go–no-go microchoices' before choosing to fully enter a target arm. This switching to and fro is very similar to that observed by Cohen and Eichenbaum in the simultaneous olfactory discrimination case.

A similar effect was found by Warrington and Weiskrantz (1974) with human amnesics, using an experimental procedure (discussed earlier in this chapter) designed to resemble the reversal learning paradigm which has been so consistently sensitive to hippocampal damage in animals (Appendix 8). Recall was cued by sets of three letters which could be the initial letters of only two English words. Subjects first learned a list consisting of one word from each of these pairs of alternatives; they were then reversed to the

alternative list. Amnesics were not significantly different from controls in their retention of the first list, but they were severely impaired both on reversal to the alternative list and on re-reversal back to the original one. Furthermore, this impairment was associated with an abnormal persistence in producing items from the first list learned.

In all of these cases, then, the critical aspect of interference in relation to amnesic performance appears not to be a degradation of initial information or increased decay of the memory trace. Rather, it consists in a failure on the part of the lesioned animals or human patients to engage in the same active suppression of competing incorrect responses as unlesioned subjects. Where the correct response is not in competition with a prepotent incorrect response, then normal (or even superior) performance can be obtained.

8.15 THE ROLE OF THE HIPPOCAMPAL FORMATION IN MEMORY

We have compressed rather a large body of data into a single chapter, mostly by the device of using a number of theoretical positions to act as summaries of substantial parts of these data. These summaries leave us with a number of pragmatic requirements for any theory of the role of the hippocampal formation in memory.

Any theory must account for the roles of time, task complexity, space, working memory, interference, etc. in determining the sensitivity of tasks to hippocampal formation damage. Further, it seems well established that the hippocampal formation is not the final repository for any long-term memories. While it remains possible that it is an intermediate repository for some restricted types of memories, there is no clear indication as to the type of memory for which it could be even the intermediate repository. For all of the 'types of memory' which we have considered, there are clear examples (as emphasized by Rawlins 1985) which are not sensitive to hippocampal formation damage. It is in any case extremely difficult to envisage a means of holding and transferring intermediate memories in a manner consistent with the known facts of brain architecture and operation.

This seems rather a tough assignment for any memory theory, perhaps accounting for the enormous number of extant theories and variants of them. We believe, however, that the problems arise from attempting to encompass the results exclusively within a memorial perspective in the first place. We have good evidence that hippocampal lesions affect tasks which are intrinsically non-memorial (Appendix 8). Rather than excluding these from consideration, in the hope of simplifying the memory problem, let us include them in the hope of solving this as a special case of a more general problem. As indicated above, our proposed solution is to view the hippocampal deficit as unitary, the same in simple conflict situations as in memorial tasks. Specifically, hippocampal damage removes the capacity to eliminate strongly competing, conflicting response alternatives as the result of a loss of the capacity to amplify the negative associations of the diverse stimuli that lead to those alternatives, or the negative associations of response alternatives concurrently primed by some single stimulus. In the case of an approach–avoidance conflict, such damage weakens the avoidance component of the conflict. In the case of a memory task, it weakens the capacity to eliminate competition from incorrect

alternatives for recognition or recall. Given a pandemonium model of memory retrieval, the hippocampal damaged subject chooses too quickly among the available goals (Meck 1988; Cohen and Eichenbaum 1993, Fig. 6.2) and hence chooses incorrectly.[2]

The strength of this hypothesis resides in the fact that it essentially conflates many of the previous hypotheses and is able to account for the data upon which they variously rested. Its weakness is that it often requires us to know special species-unique, and indeed situation-unique, details of the experimental design or of the animal's specific solution to a problem, details not necessarily mentioned in the methods or results sections of the papers concerned. However, this weakness in applying the theory to the data currently available does not entail that the theory is not testable with future experiments.

2. With verbal and conscious retrieval the result of a pre-verbal or pre-conscious pandemonium will often be that no apparent retrieval occurs at all.

9 Fundamentals of the septo-hippocampal system

So far, we have sketched out a psychological view of a behavioural inhibition system, argued that anxiolytic drugs act upon it, and used drug–lesion comparisons to derive the conclusion that anxiolytic drugs are more likely to act by impairing septo-hippocampal function than by impairing function in any other structure in the brain. To take these ideas any further we must have a detailed view of how the septo-hippocampal system works and what specific effects the anxiolytic drugs have on it. Our efforts will lead us not just to a theory of the psychology of anxiety or of its neurology or pharmacology, but to an integrated neuropsychology of anxiety which will blur (as blurred it must one day be) the division which at present separates our two languages, of brain process and mind.

This blurring already appears to be occurring in those recent computer models which attempt to have both neurally realistic architecture and neurologically and psychologically testable output. The marriage of psychology and neurology via such cybernetic models forces out many of the hidden false assumptions buried within each discipline (e.g. Hinton *et al.* 1993). There are indications that analysis of the septo-hippocampal system itself may have reached the stage where such models are profitable (e.g. *Hippocampus* 1996, Vol. 6, Issue 6). However, the lack of many critical data (such as the nature of the subcortical inputs to the theta control system) suggests to us that such modelling for the septo-hippocampal system as a whole would be premature. Nonetheless there are many data and many types of data which are available and we have seen it as our job to produce a theory of the septo-hippocampal system and of anxiolytic drug action which attempts (but at a much less detailed level than that of a computer model) to accommodate all the available data. This means that we had to adjust conclusions drawn from clinical and experimental psychology to conclusions drawn from pharmacology and physiology, and vice versa. Similarly, the functions we wished to ascribe to the hippocampus had to be made consistent with its known connections, with the functions ascribed to the structures to which it is connected, and with the specific nature of the partial interference with its function produced by the anxiolytic drugs.

As a background to understanding anxiolytic drug action, and hence anxiety, we therefore present below the key conclusions we have drawn about the septo-hippocampal system (these are more fully justified in the appendices). We then derive from them a set of principles. These principles are often phrased in a way which will make them particularly easy to apply to the theory presented in Chapter 10. Their content is, nonetheless, tightly constrained by our detailed data reviews and so they can be taken as principles with which any current theory of septo-hippocampal function or anxiolytic action should comply. This will lay the groundwork for the theoretical integration of the next two chapters. These provide not only a view of the role of the hippocampus

in memory (largely derived from the conclusions of Chapter 8) but an integrated theory of all the behavioural functions of the septo-hippocampal system (Chapter 10) and a neuropsychology, and consequently typology, of clinical anxiety disorders (Chapter 11). In order to follow a coherent argument, some key points from previous chapters will be briefly reiterated.

9.1 ANXIOLYTICS AND THE SEPTO-HIPPOCAMPAL SYSTEM—AN OVERVIEW

The key to our analysis of the septo-hippocampal system (as with the rest of the book) is one major premiss: that anxiolytic drugs can indicate at least part of the neurology of anxiety. It follows that a description of a psychological process that is altered by both classical and novel anxiolytics is extremely likely to be a description of part of the psychology of anxiety; and a description of a neural process altered by both novel and classical anxiolytics is likely to be a description of part of the neurology of anxiety.

9.1.1 **Anxiolytic action on the septo-hippocampal system**

Our use of this major premiss causes us to emphasize in our theoretical scheme many data which others would simply ignore. This is because the most important single fact for this book is that all known clinically effective centrally acting anxiolytics produce the same characteristic changes in the functioning of the septo-hippocampal system, and these changes are not duplicated by any drugs that have not been shown to be anxiolytic in the clinic. Moreover, the anxiolytic drugs produce these common changes through a wide number of pharmacologically and anatomically different routes. These are objective data. Whether or not you share our particular perspective, they must be accounted for by any theory of hippocampal function which pretends to be complete.

The anxiolytic drugs share not one but two actions on theta activity in the septo-hippocampal system. To remind the reader, this activity consists in a synchronous burst firing of cells which is thought to be an important functional state of the hippocampus and related brain structures, and which can give rise to a gross extracellularly recorded rhythm. The frequency of this activity is controlled by impulses coming from a 'pacemaker' in the medial septum (Appendix 5), and stimulation of the septum at theta frequencies can produce a rhythm, indistinguishable from spontaneous activity but phase locked to the stimulus (Appendix 5; see Fig. 9.1A). The septum, in turn, receives input from hypothalamic nuclei in which the frequency of theta is determined (Appendix 5).

The first action of anxiolytics on theta to be reported, and the one on which the theory of the first edition was primarily based, was an action on septal driving of hippocampal theta. As shown in Fig. 9.1C, anxiolytics increase the threshold for the septal driving of hippocampal theta rhythm (Appendix 5). This increase occurs predominantly in the middle of the frequency range, and can be reproduced by specific lesions of the noradrenergic input to the hippocampus from the locus coeruleus.

The second action of anxiolytics on theta to be reported was an action on the frequency of reticularly elicited theta. As shown in Fig. 9.2, anxiolytics reduce this reticularly elicited

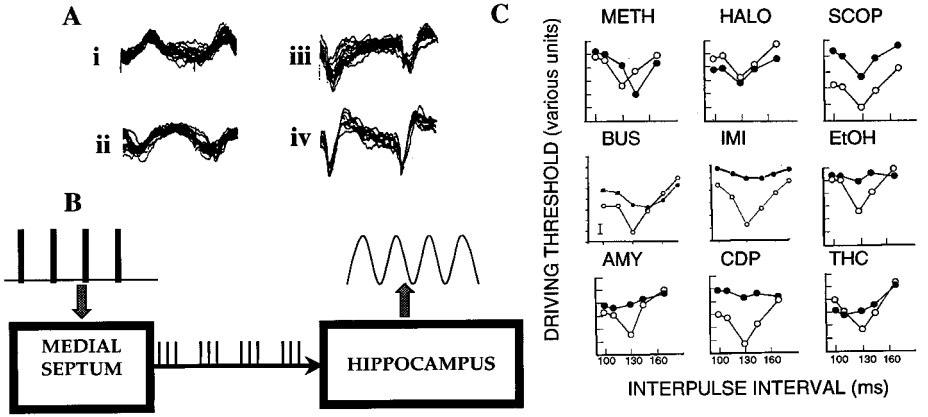


Fig. 9.1 Septal driving of theta rhythm—common effects of anxiolytic drugs. (A) Trains of low frequency (5–12 Hz) stimulation delivered to the medial septum (see B) can induce synchronous activation of the hippocampal formation (here the dorsomedial subiculum). Each panel is the superimposition of about 10 successive sweeps with two stimuli from the train within each sweep (see small sharp descending stimulus artefacts in i). The responses shown in i and ii are examples of theta driving in two different rats: the synchronous responses are sinusoidal and have the same conformation as spontaneous theta waves of the same frequency (7.7 Hz) in that rat. ii and iv are driven theta and non-theta evoked potentials obtained in two animals which only showed driving or evoked potentials respectively. i and iii show a shift from driving to evoked potentials within a single rat with an increase in stimulation intensity. To detect the effects of anxiolytic drugs it is important that the stimulation should only produce theta driving. (Waveforms from James *et al.* 1977.) (B) Diagrammatic representation of theta driving. The imposed electrical pulses activate the septum and cause it to produce the same bursts of impulses (shown between the septum and hippocampus) as are produced in the generation of spontaneous theta activity. These entrain similar activity in the hippocampus which, because it is synchronous in many cells, can give rise to a high-voltage (up to about 1 mV) sinusoidal wave form—theta rhythm. Note that each pulse, burst of septal firing, and theta wave is synchronous (with a slightly increasing phase lag due to conduction delays). (C) The effects of anxiolytic drugs on the threshold for driving theta rhythm. Septal driving stimulation (see panels A and B) produces theta rhythm at a threshold which varies in a U-shaped fashion with frequency in male rats (open circles; James *et al.* 1977; see Drewett *et al.* 1977 on female rats). Treatment with anxiolytic drugs (BUS, IMI, EtOH, AMY, CDP, THC) increases the threshold for driving (filled circles), particularly in the middle of the frequency range (McNaughton *et al.* 1977; Zhu and McNaughton 1994). Treatment with drugs that are not anxiolytic does not usually produce this effect although they can alter the location of the minimum of the curve (METH, see also McNaughton *et al.* 1980a,b; Valero *et al.* 1977) or reduce thresholds generally (HALO, see McNaughton *et al.* 1977). Treatments which impair noradrenergic function have similar effects to anxiolytics (Gray *et al.* 1975). Scopolamine (SCOP) does not affect thresholds as such. Rather, as it does with spontaneous and reticular-elicited theta, it totally blocks theta that is not accompanied by movement. As a result it increases the average threshold for driving. METH, methysergide; HALO, haloperidol; SCOP, scopolamine; BUS, buspirone; IMI, imipramine; EtOH, ethanol; AMY, amylobarbitone; CDP, chlordiazepoxide; THC, Δ^9-tetrahydrocannabinol.

frequency (and they also reduce spontaneous frequency while the rat engages in a range of behavioural tests). This frequency change is not specific to any part of the frequency range, and cannot be reproduced by depletion of noradrenaline. It can, however,

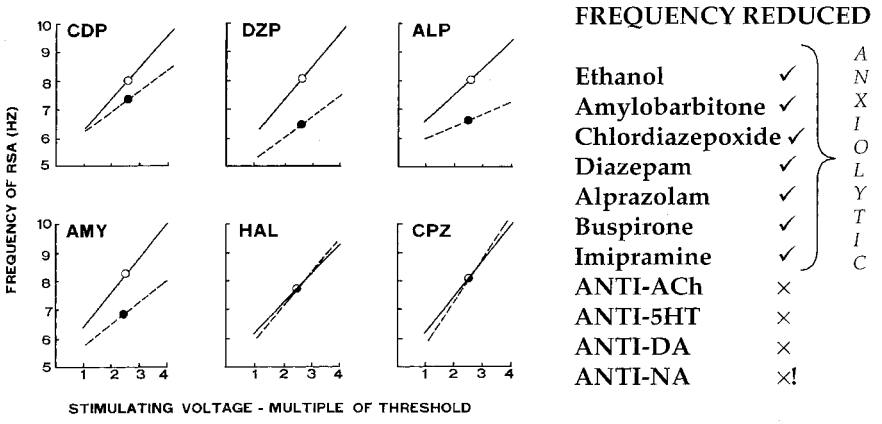


Fig. 9.2 Reticular-elicited theta rhythm—common effects of anxiolytic drugs. Left: increasing levels of reticular stimulation produce a linear increase in the frequency of theta rhythm (open circles). Anxiolytic drugs (CDP, DZP, ALP, AMY) decrease the frequency of theta (filled circles) while sedative but non-anxiolytic drugs (HAL, CPZ) do not. (From McNaughton *et al.* 1986.) Right: this effect is produced by all known classes of clinically effective anxiolytic drug and is not the result of reduced activity in cholinergic, serotonergic, dopaminergic, or noradrenergic systems (McNaughton and Sedgwick 1978; Coop *et al.* 1990; McNaughton and Coop 1991). CDP, chlordiazepoxide; DZP, diazepam; ALP, alprazolam; AMY, amylobarbitone; HAL, haloperidol; CPZ, chlorpromazine; ACh, acetylcholine; 5HT, 5-hydroxytryptamine; DA, dopamine; NA, noradrenaline.

be reproduced by injections of anxiolytic benzodiazepines into the supramammillary nucleus (McNaughton *et al.* 1995) and the dorsomedial nucleus of the hypothalamus (McNaughton *et al.*, unpublished observations).

Thus, anxiolytic drugs are unique in impairing these two independent aspects of the control of hippocampal theta rhythm. Conversely, these two tests are essentially unique in demonstrating similar dose-related physiological effects of the different classes of anxiolytic drug. These data strongly suggest, therefore, that anxiolytic drugs achieve their clinical effects by impairing the control of theta rhythm in the hippocampus and related structures such as the entorhinal cortex and posterior cingulate cortex (Appendix 5).

9.1.2 **Anxiolytics share the behavioural effects of septo-hippocampal lesions**

If theta activity is important for hippocampal function, the impairment of theta control produced by anxiolytic drugs should make them act like weak and/or selective hippocampal lesions. This prediction holds true over a surprisingly wide range of behaviour (Chapter 4, Table 4.2), given that the drugs do not eliminate theta, far less eliminate the whole hippocampus.

Of particular note is the fact that anxiolytic drugs impair behaviour in tasks such as delayed matching-to-sample and spatial navigation which are often thought of as being particularly 'memorial' or 'hippocampal' tasks. Of equal note is the fact that the effects of hippocampal lesions include retarded extinction without change in acquisition of simple

running in an alley-way, and include impaired passive but not one-way active avoidance. These results show that hippocampal lesions do not affect memory in any simple way; indeed, we argued in Chapter 8 that they do not affect memory directly at all.

9.1.3 **Changes in theta control mimic anxiolytic action and hippocampal damage**

The fact that anxiolytics impair hippocampal function and that they produce behavioural effects like hippocampal lesions provides only correlational evidence for the view that the drugs produce their behavioural effects because of their action on theta. A direct test of the role of the anxiolytic change in septal driving of theta is provided by investigation of the behavioural effects of noradrenergic depletion (which reproduce the first effect of the anxiolytics on hippocampal electrical activity noted above). As can be seen from Table 4.2, specific neurotoxic lesions of the dorsal ascending noradrenergic bundle reproduce in addition a large part of the common behavioural profile of anxiolytic drugs and septo-hippocampal lesions (see also Appendix 10). Similarly, a direct test of the role of the anxiolytic reduction in theta frequency (the second effect noted above) is provided by investigation of the behavioural effects of injections of benzodiazepines into (and neurotoxic lesions of) the supramammillary nucleus and the dorsomedial hypothalamic nucleus. Such studies have been carried out only in the last year or so, but the evidence so far (Appendix 6) is that, together with dorsal bundle lesions, they could account for the whole common behavioural profile of anxiolytic drugs and septo-hippocampal lesions. Particularly in tests insensitive to dorsal bundle lesions but involving behavioural inhibition, injections of benzodiazepines into the supramammillary nucleus can have effects on theta frequency and on behaviour which are as large as those of systemic injections of the same drug (Senior *et al.*, in preparation; for details, see Appendix 6). A further way to test the causal role played by hippocampal theta in mediating the effects of the anxiolytic drugs is by investigating the behavioural effects of direct manipulation of the theta rhythm, using septal driving stimulation. Experiments of this kind, focusing on resistance to extinction of both running and lever-pressing responses, have shown that such 'theta driving' at frequencies close to 7.7 Hz (the frequency at which systemic anxiolytic administration maximally elevates the driving threshold) produces, as predicted, behavioural effects opposite in sign to those of systemic administration of the drugs themselves (Williams *et al.* 1989; Appendix 7).

So, when the characteristic changes in hippocampal electrophysiology produced by anxiolytic drugs are reproduced by highly selective intracranial manipulations, these appear to reproduce the bulk of the behavioural profile of systemic anxiolytics. Conversely, when electrophysiological changes of opposite sign to those caused by the drugs are induced experimentally, so the behavioural profile, too, is opposite in sign. This leads us to

Principle 1. The septo-hippocampal system mediates anxiolytic drug action and hence at least some aspects of anxiety.

However, in interpreting this principle we should note that not only are the behavioural effects of the drugs sometimes mediated by changes in the threshold control of theta (as evidenced by the effects of dorsal bundle lesions) and sometimes by changes

in theta frequency (as evidenced by experiments on the supramammillary nucleus and the dorsomedial hypothalamic nucleus), but also behaviourally significant changes in frequency may be mediated by different systems under different behavioural conditions (Fig. 9.3).

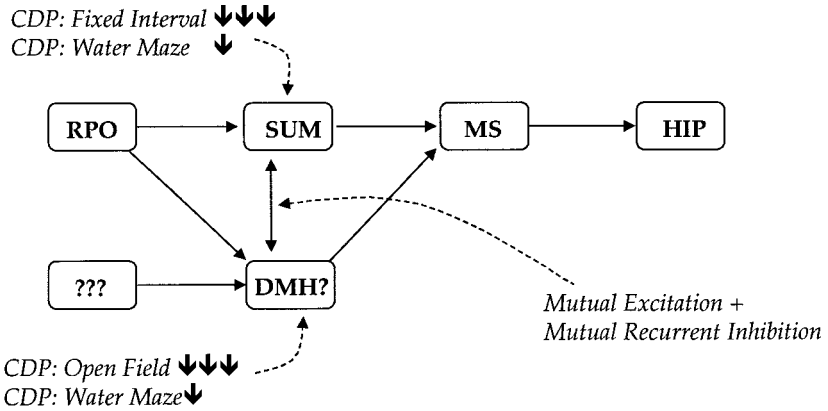


Fig. 9.3 The neural control of theta frequency. Electrical stimulation of the nucleus reticularis pontis oralis (RPO) activates the medial supramammillary nucleus (SUM) which encodes theta frequency (probably via recurrent inhibitory GABAergic interneurons). This is relayed to the hippocampal formation including entorhinal and posterior cingulate cortices (HIP) by the medial septum–diagonal band complex (MS). While animals work on a fixed interval task, injections of chlordiazepoxide (CDP; anxiolytic benzodiazepine, net GABA-agonist) into SUM reduce theta frequency and change behaviour as much as does systemic injection. In the water maze, CDP in SUM has only weak effects on theta frequency and spatial navigation. In the open field, unlike systemic CDP, CDP in SUM has no effect on theta frequency or behaviour. We know that lesions of RPO and lesions of SUM do not eliminate theta rhythm, but that lesions of SUM reduce the frequency of theta rhythm. It follows that there must be a number of areas (including RPO) which can activate a number of benzodiazepine-sensitive, frequency-controlling nuclei (including SUM and probably dorsomedial hypothalamus (DMH)). The latter must integrate their inputs via mutual excitation and phase lock their outputs via mutual inhibition. Different activating and frequency controlling nuclei appear to be involved to different extents in different behavioural tasks. Arrows: ↓ ↓ ↓, large impairment; ↓, small impairment.

9.2 THE DEFENCE SYSTEM

We started this book with an analysis of the hierarchical organization of the defence system. This choice was driven by our interest in anxiety. However, the defence system is a particularly well-understood exemplar of many motor control systems with which the septo-hippocampal system has intimate connections and whose functions constrain the functions that can be imputed to the hippocampus. It provides, therefore, a convenient starting point for functional analysis of the septo-hippocampal system.

In Chapter 2, we discussed the ethology of threat and concluded, following the work of Robert and Caroline Blanchard, that immediate, present threat elicits reactions

fundamentally different from those elicited by anticipated threat. In each case reactions also vary depending on the effective 'defensive distance'. Chapters 3 and 4 allowed us to conclude that anxiety, at least as defined by the effects of the anxiolytic drugs, results from anticipated rather than immediate threat, and that it produces both behavioural inhibition and increases in risk analysis, including active risk analytic behaviour patterns.

An important shift in emphasis from the first edition is our highlighting that pure fear (when active escape or avoidance is all that is required) and anxiety (when fear, or some other aversive state, is compounded with a requirement to approach the source of aversion) can each be either conditioned or unconditioned. Thus, where the first edition treated innate anxiety-provoking stimuli and novelty as special cases which required separate justification for their inclusion as inputs to the behavioural inhibition system, the present, more ethological approach includes them automatically. To make this integration possible, we added to the general notion of defensive distance the independent general notion of defensive direction: that the critical feature engaging anxiety systems is the requirement to approach (as opposed to escape from) a source of danger (or potential danger), thus giving rise to an approach–avoidance conflict.

Chapter 6 reviews a coherent neurology (borrowed largely from LeDoux and Graeff) which maps onto this ethological and behavioural analysis (shown in Fig. 1.8). This neurology starts with the dorsal periaqueductal grey as a key centre for the control of immediate responses to a predator: freezing, fight, flight, autonomic output, and analgesia. Specific lower subsystems controlling these different reactions are also known, but differentiation between them is not necessary for our present purposes. The dorsal periaqueductal grey is activated by an immediate predator, by an immediate dominant conspecific, by pain, and by high levels of carbon dioxide. Graeff (1994) has suggested that activation of this region is the basis of panic attacks, and we accept this suggestion (with a few minor additions later). It is important to note that the flight (escape) elicited by activation of the periaqueductal grey is undirected.

The periaqueductal grey is connected to the medial hypothalamus, where more sophisticated escape mechanisms appear to be located. This in turn is connected to the amygdala, which coordinates simple avoidance. In LeDoux's formulation the amygdala is then viewed as being connected to the hippocampus, which mediates complex, particularly inhibitory, avoidance. However, we argue (Chapter 6, Appendix 3) that the anterior cingulate represents several additional levels of the active defence system and hence is directly above the amygdala in an 'active defence' hierarchy. By contrast, the hippocampus and, above it, the posterior cingulate appear to be parts of a parallel inhibitory system principally concerned with the approach into dangerous situations (Fig. 1.8).

It should be emphasized that all these areas are reciprocally interconnected, but that the connection from the amygdala to the hippocampus appears to be feedforward, while that from the hippocampus to the amygdala appears to be weaker and feedback (Appendix 2). Further, a review of the monoaminergic systems (Appendix 10) suggests that these regulate the interactions between the components of the defence system, so that, for example, the amygdala and periaqueductal grey are affected inversely by the same serotonergic input (Fig. 6.4). This makes adaptive sense, since the more you know about how to avoid a threat efficiently, the less useful is an undirected escape reaction.

However, the organization of the sensory inputs to the defence system raised an interesting question. Is the hippocampus, rather than the anterior cingulate, the top of the 'motor' threat system (receiving sensory input as do the amygdala and periaqueductal grey, and controlling responses when these structures do not), or is it the top of the 'sensory' threat system (sending a particularly highly organized type of sensory information to the amygdala, which then controls the response to threat)? Both points of view appear to be correct (LeDoux 1994). The hippocampus needs to be engaged in the organization of avoidance responses when these conflict with other classes of goal (e.g. approach) and hence require inhibition of a prepotent response. However, the hippocampus is also required when there is competition between incompatible specific goals (stimulus sources) to which to direct the same appropriate avoidance response.

An important corollary of LeDoux's formulation is that each of the nuclei in the hierarchical threat system (from the periaqueductal grey to the anterior cingulate) must be presumed to be receiving sensory information at all times. Hence, given adequate stimuli, all of these nuclei will be concurrently pre-programming a range of quite independent potential responses. This multiplicity of potential responses may seem complicated, but it can easily be controlled by a system which depends on release (inhibition of inhibition), rather than activation, to select responses (see, for example, McNaughton 1989a, Chapter 2).

Thus, if a more sophisticated level of the system such as the amygdala is highly active, this should normally inhibit the 'quicker and dirtier' responses which would otherwise be the output from a more basic level of the system (such as the periaqueductal grey). The concurrent high-intensity activation of multiple networks, coupled with control by release from inhibition, allows the sudden dramatic shifts which can be seen between, say, freezing and defensive attack. This pattern of behaviour can be attributed to a shift of disinhibitory influences from, for example, a (continuously updated) freezing motor program to a (continuously updated) defensive attack motor program. Note that this should not normally result in what we would term goal conflict. The goal is the same for the two systems (i.e. achievement of a place of safety); it is only the selection of responses to achieve that goal which must be controlled, and this can be achieved within the hierarchical system without the need for outside influence.

As with modern network models of many processes, all of these structures are reciprocally connected. Despite our description of, for example, the periaqueductal grey as controlling escape, it is probably much better to view this as part of a distributed system, including the hypothalamus, amygdala, anterior cingulate, posterior cingulate, and hippocampus. Certain patterns of activity in this network cause output from the periaqueductal grey to engage the various fight–flight–freeze nuclei. This region may then be the most important node through which information passes during fight, flight, and freezing; but it would be an error to see it as the only structure in the threat system active at that time, or to think that it is the only structure in the threat system which is contributing to escape.

Our analysis of the threat system has, so far, closely followed those of LeDoux and Graeff (see Chapter 6 and Appendix 2). However, we differ from them in thinking that the hippocampus should not be viewed as a source of contextual stimuli which can come to control the amygdala (for then there should be strong hippocampo-amygdalar connections rather than, as is the case, the other way round). Rather, we think of the hippocampus as a device which can resolve conflicts, not only between programs

concurrently activated within the amygdala, but also between programs in the amygdala and those elsewhere. Thus, the hippocampus does appear to be important in dealing with some cases where contextual stimuli determine defensive behaviour. However, *ex hypothesi,* the role of the hippocampus will be important only when the contextual stimuli elicit conflict. Equally, the hippocampus is also important for the way in which the amygdala deals with non-contextual stimuli. It will be involved in any case, for example, where the stimuli generate approach–avoidance conflict. While contextual stimuli, being diffuse, will often have this effect, this is a correlational rather than a necessary relation. That is why, on our view, contextual or configural views of the hippocampus can appear to provide satisfactory accounts of many of the hippocampal effects in memory-oriented experiments, but do not account for all such effects and lacks of effect (Chapter 8), and do not account for the non-memorial effects of hippocampal lesions.

This analysis of the hierarchical threat system is very close to what is known of the much simpler eye-blink conditioning system (Appendix 6). The latter appears to be a paradigmatic case of a hierarchically organized system. The lowest level (nucleus interpositus) is all that is required for the simplest conditioning, but is also the final common path for more complex conditioning. In this system the cerebellar cortex stands in about the same relation to interpositus as do hypothalamus and amygdala to the dorsal periaqueductal grey, providing a capacity for more anticipatory responses. Finally, occupying the same logical position in both systems is the hippocampus. This provides the higher level eye-blink circuits with the additional capacity for inhibition of the previously correct response during reversal, and for inhibition of multiple incorrect alternatives provided by multiple cortical delay lines during trace conditioning (Appendix 6).

Thus, superimposed on the most basic conditioning circuits are higher-order stimulus circuits (which can provide simple output from detectors of progressively more complex stimulus configurations), including those which can deal with delays and relations between stimuli. Despite the complexity of the networks involved, all of the resultant conditioning is, from the point of view of the target structure, identical to that of the simplest possible stimulus input; all that differs is the source of the afferent impulses, and the receiving cell has no way of telling what that is. Depending on the precise conditions of the experiment, this multiple parallel circuitry will on occasion tend to produce concurrent activation of two incompatible goals. It is the business of the hippocampal system to detect this conflict between proposed goals and send out signals which are designed to devalue all but the correct one.

The clear parallels between these two 'simple conditioning' systems (fear and eye-blink) lead to our second important principle.

Principle 2. Conditioning, as such, reflects simple changes in associative strength at single synaptic junctions in relevant systems distributed throughout the brain (amygdala for fear conditioning, interpositus for eye-blink). These changes can take the form of increases as a result of long-term potentiation (or equivalent plastic changes) or decreases as a result of long-term depression. Long-term potentiation and depression in the hippocampus, on the other hand, do not represent conditioning (or storage of sensory information), but the reprogramming of a system which modulates conditioning (or storage of sensory information) in other structures.

9.3 THE MOTOR/WORKING MEMORY SYSTEM

If we view motor cortex as an area where specific movements are organized, then the supplementary motor area (to which it is reciprocally connected) is where anticipated movements (plans) are organized. This is in turn reciprocally connected to the prefrontal cortex, which is hierarchically organized with reciprocal connections between levels (Fig. 1.9). Each level of this hierarchy (working back from motor cortex) appears (Appendix 3) to organize an aspect of motor performance one step more anticipatory than the previous one. Thus, the frontal eye fields represent a system which controls the direction of gaze so that the direction of the next motor act can then be computed. Similarly, the area in the region of the principal sulcus can be thought of as controlling the direction of 'gaze' into the memory stores of the posterior cortex, so that, for example, future eye movements can be computed.

This same hierarchical organization is seen, in very close parallel, in the separate dorsal ('where') and ventral ('what') trends in frontal cortex. These are reciprocally connected with each other, thus binding the what and where information together.

The prefrontal cortex as a whole can be seen both as providing high-level (essentially inhibitory) control of the ordering of ongoing motor programs, and as having mapped into it a variety of types of working memory. What distinguishes this region from areas such as the amygdala or cingulate is the high degree of flexibility of the resultant motor output. Given the differentiation already demonstrated in prefrontal cortex, it is possible that different types of working memory are uniformly topographically mapped into different parts of prefrontal cortex. Thus visuo-spatial information is transferred from posterior parietal cortex in direct projections to the principal sulcus, where it provides the basis for visuo-spatial working memory; and object attributes are transferred from visual association areas to below the principal sulcus, where they provide the basis for visual working memory.

At first sight, motor programming and working memory might seem quite distinct entities. But, in line with McKay's (1987) equation of perception with action, the information being held in working memory effectively defines the way motor programs will be controlled. This is particularly the case if the working memory holds temporary versions of motor programs recalled from long-term storage, as well as temporary versions of stimulus items. In one of the standard tasks used to test working memory in monkeys, where the monkey must move its eyes to a new position after a delay, it should be noted that the 'memory of the stimulus' is *identical* to the 'goal of the upcoming motor act'. This intertwining of stimulus and response aspects is something which, in relation to the hippocampus, has caused many attempts at pure stimulus or pure response theories to come to grief, and constitutes one of the reasons we have preferred the term 'goal' to describe the nature of the encoded information.

Principle 3. At the level of operation of the prefrontal cortex and the septo-hippocampal system no distinction should be made between 'stimulus' and 'response'. Both should be conflated into a term such as 'goal' (hippocampus) or 'plan', i.e. the ordering of successive goals and subgoals (prefrontal cortex).

Theoretically, all that is required for the use of working memory is recursive 'refreshing' of activity by reciprocal connections between some area of prefrontal cortex ('working') and the equivalent area of more posterior cortex ('memory'). In many cases, sensory input will take the form that only one such circuit will control behaviour. However, as with the threat system, there exist multiple possibilities for the coactivation of different prefrontal–posterior cortical networks so as to produce different working memories.

As with the hierarchical ordering of the threat system, the hierarchical organization of frontal cortex (in terms of progressively higher levels of anticipation) can normally provide a means of resolving goal–subgoal conflicts or, which is computationally the same thing, organizing sequences of goals in time. However, at any individual step in a complex motor sequence there may be uncertainty as to which of two goals is correct. (This would be particularly true in, for example, a working memory task with a high level of interference from earlier trials.) Uncertainty of this kind would involve a conflict between areas coding for different goals within a specific level of prefrontal cortex (i.e. at the same nominal step in the upcoming action sequence) rather than between levels of the system (i.e. at different steps in the upcoming action sequence). We propose that, when two or more such working memory circuits are highly and equally active at one time, the hippocampal formation can become engaged in order to resolve the conflict between the incompatible upcoming goals represented by the appropriate working memory systems.

The prefrontal system has major links to the perirhinal and parahippocampal cortices, links that provide the main basis for its interactions with the hippocampal formation. There may be cases when only the entorhinal, perirhinal, and parahippocampal cortices interact with prefrontal cortex (as, perhaps, with concurrent discriminations; see Appendix 9, Section A9.3). These purely cortical interactions are likely to involve conflicts between goals, the distinction between which is related more to the stimulus to be acted on than to the specific actions to be taken with respect to the stimulus (see G3 and G4 in Fig. 1.7, p. 28): conflict, in essence, between memory items. But there are also cases where the hippocampus proper is involved. These appear to consist in conflicts between goals which differ more in the action to be taken than in the stimulus to be acted on (see G2 and G3 in Fig. 1.7), although the relevant stimuli may also be quite different, e.g. food in one location versus safety in another.

On this view, the recursive connections of the hippocampus and the frontal cortex would allow the hippocampus to do two things. First, the hippocampus could alter processing in the working memory system (for example producing an increase in negative bias of the type ascribed to generalized anxiety disorder by Eysenck 1992a), and hence affect motor programming indirectly. Second, it could actively interrupt ongoing motor programs and replace them with specific information gathering (e.g. risk analysis) programs. The same logic can be applied equally well to higher-order, more anticipatory, computations (planning) and lower-order, less anticipatory, computations (attention, gaze direction).

Here it will be seen that our view of frontal cortex follows closely those of Goldman-Rakic, Fuster, Pandya, and Barbas, as well as many others (Appendix 3). However, as with our view of the threat system, what generally distinguishes our theory are the ideas: (1) that all memory is located outside the hippocampus; and (2) that the hippocampus

is a quite specialized structure which acts solely to resolve conflicts between incompatible goals. Conflicts of this kind are fundamentally linked to response programming. However, the case in which working memory holds a goal in respect of which the critical information is predominantly stimulus-related (determining, for example, towards what object to direct the eyes next) shows that the distinction between response competition and memory competition may be hard to make; and there is nothing about the machinery we attribute to the wider hippocampal formation, including the entorhinal cortex, which would prevent its resolving a pure stimulus competition, in the sense of a conflict between two goals both of which can be acted upon by what is fundamentally the same response. Certainly, the hippocampal formation does receive some purely stimulus information, which, however, it appears to use predominantly for response control purposes.

In addition to the resolution of conflicts between concurrently activated goals (dealt with by the hippocampus), a task may require resolution of conflicts between different subgoals contributing to an overall goal (essentially the problem of ordering the different subgoals into an appropriate sequence). We see this latter resolution as the business (perhaps, in essence, the only business) of the prefrontal cortex. Thus, the hippocampus and prefrontal cortex deal with different kinds of behavioural inhibition. In some cases (e.g. delay tasks), however, it is difficult to distinguish between opposing response requirements (to respond on the one hand and withhold responding on the other) and sequencing (withholding responding now and responding later). In such tasks performance could depend on both the hippocampal formation and the prefrontal cortex.

Note also that frontal cortical systems are in parallel to the defence and eye-blink systems, in the same sense that each of these is in parallel to the other. Each can be concurrently and separately active, with each determining some course of action on the basis of the information available to it. The eye-blink system is the most fundamentally reflexive, the defence system is to a large extent based on fixed action patterns (in the most flexible sense of this term), while the frontal cortex (with primary and secondary motor areas viewed as the lowest levels of the system) appears to deal with the least reflexive of motor outputs.

Principle 4. Motor control, attention, working memory, and planning reflect successively deeper levels of processing in otherwise fundamentally similar processing units located in the frontal cortex. The hippocampus has no *necessary* involvement in any of these functions. Rather, it operates to modulate the activity in frontal circuits only under conditions where *concurrent* goals conflict.

9.4 THE 'EMOTION' SYSTEM

Following Papez, and by direct analogy with classical sensory systems, it is tempting to view the anterior cingulate gyrus as being emotional sensory cortex (receiving its input fairly directly from, for example, the pain system). By analogy with other sensory systems, its reciprocal links with the prefrontal cortex would then be viewed as supporting emotional working memory. However, the input to the anterior cingulate gyrus from

the amygdala and hypothalamus, and its immediate links with motor cortex and dorsal and ventral striatum, speak of a motor programming role, making it akin to a second ventral prefrontal cortex.

In the case of the classical perceptual systems, we found that we could equate perception with action only at the very highest levels of these systems, at the point where they are involved in working memory. However, this equation is much easier to make in the case of emotional perception, since the detection of an emotional state of affairs can reasonably be presumed to imply immediate action of a quite specific type. There is also evidence that cingulate cortex is not primary pain-receiving cortex, but instead is concerned with the organization of responses to pain. Cingulate lesions in human beings do not eliminate the capacity to detect pain and report changes in its level, but they do rob it of its emotional, imperative quality.

Let us consider further the fact that the cingulate cortex receives information not only from the hypothalamus (which can be presumed to relate to innate motor programs activated by current internal states), but also from the caudate nucleus and amygdala. According to Rapoport (1989), these pathways allow information about innate motor programs (grooming, checking territory, etc.) to gain access to the frontal cortex. She proposes that, in at least some cases of obsessive–compulsive disorder, an abnormality in the basal ganglia (of a type which gives rise to choreas and tics in a variety of other disorders) releases such innate programs. The resultant activity in cingulate and frontal cortices then gives rise to the compulsions and obsessions, these being perceived in consequence as 'intrusive and senseless' (Rapoport 1989, p. 63). This line of thought leads to two, mutually compatible, ways of regarding obsessive–compulsive disorder. First, one can treat the symptoms of this disorder as reflecting activity in the highest goal-setting levels of the motor programming system. On this view, these symptoms consist of affective action programs repetitively released despite stimulus conditions which should normally inhibit them. Second, one can treat obsessive–compulsive symptoms as reflecting activity in the highest levels of the affective perceptual systems. On this view, they consist of affective percepts occurring in the absence of a normally adequate stimulus and hence giving rise to inappropriate behaviour which, because it cannot correct a condition that does not objectively exist, is remorselessly repeated. Since, at this level of the nervous system, the distinction between action and stimulus is, as we have emphasized, arbitrary, these are two different ways of saying exactly the same thing.

In this light, anterior cingulate is concerned with the control of relatively innate motor programs which each fulfil phylogenetically distinct functions. Hence it has strong links with the amygdala, dorsal and ventral striatum, and motor cortex. It functions, therefore, very much like ventral frontal cortex, except that the latter is concerned with the production of flexible motor programs which have little innateness.

What of posterior cingulate cortex? This appears (Appendix 3) to be a somewhat ambiguous structure. It is tightly and reciprocally linked to anterior cingulate, and yet stimulation does not elicit affective responses as it does in anterior cingulate. It is placed in the cortex close to the subiculum, receives largely unidirectional links from subiculum, as does subiculum from other hippocampal areas, and shows the theta activity and perhaps theta rhythm which are characteristic of the hippocampal formation. Yet it has not before been treated as simply another module of the septo-hippocampal system. We argue in

Appendix 3 that posterior cingulate is in fact tightly related to both the anterior cingulate and the septo-hippocampal system. Its reciprocal links with the anterior cingulate mirror the reciprocal links between the dorsal and ventral trends respectively of frontal cortex, with anterior cingulate representing the ventral 'what' trend and posterior cingulate, the dorsal 'where' trend. The distinction between posterior cingulate and dorsal prefrontal cortex would then be the same as the distinction already proposed between anterior cingulate and ventral prefrontal cortex: posterior cingulate cortex deals with more stereotyped, and dorsal prefrontal cortex with more flexible, motor programs. Both are concerned primarily with prediction and anticipation (hence the lack of motor output when posterior cingulate is stimulated), and with the inhibition of ongoing programs.

It is in this anticipatory and inhibitory role that we can see a link with the hippocampal formation. The posterior cingulate receives many inputs which match those of prefrontal cortex (from the anterior thalamus, the anterior cingulate, the mammillary bodies, the subiculum). However, as we argue in Appendix 3, it seems to have a particularly close relationship with rhinal cortex. Matching the architectonic mapping of posterior sensory cortical areas into prefrontal cortex, we can see an apparently similar architectonic mapping of rhinal isocortex into the most differentiated layers of posterior cingulate and of subiculum into the least differentiated (Fig. 1.9). The direction of flow of cingulate information, predominantly from least to most differentiated, matches that in frontal cortex, while the flow of hippocampal information from most differentiated to least matches sensory cortex. Thus, the projection from parahippocampal cortex to the final output stage of posterior cingulate could be a direct 'quick and dirty' route for information which can also be processed through the 'slow and sophisticated' circuitry running through all the levels of the hippocampal formation and then all the levels of the posterior cingulate to arrive at the same final destination (comparable to the direct and cortical sensory inputs to the amygdala; LeDoux 1994; Fig. 6.1).

Posterior cingulate, then, as anticipatory, inhibitory, affective, innate premotor cortex, is in receipt of information from a variety of levels of the hippocampal formation and is in direct receipt of at least part of the isocortical input to the hippocampal formation. Yet we cannot view posterior cingulate as the sole or even primary cortical recipient of hippocampal output. The subiculum projects to anterior cingulate (matching its projection to the amygdala) and both directly and via the anterior thalamus to the prefrontal cortex.

We should note here that, while we have argued strongly that a crucial site of action of the anxiolytic drugs is the hippocampal formation, prefrontal and cingulate lesions have each proved effective in dealing with drug-resistant forms of anxiety. Thus, hippocampal formation, posterior cingulate, and the dorsal trend of prefrontal cortex may be different parts of a distributed system, each dealing with different aspects of essentially the same very general class of anticipatory, inhibitory problems.

Principle 5. The dorsal trend of prefrontal cortex, the posterior cingulate, the hippocampal formation, and the basal ganglia all deal with different aspects of response inhibition. The prefrontal cortex is involved primarily in the sequencing of goals, the cingulate cortex in the achievement of a range of innate goals, the hippocampus with concurrently conflicting goals, and the basal ganglia with the selection of the means to achieve a specific goal.

9.5 GENERAL APPROACH TO THE SEPTO-HIPPOCAMPAL SYSTEM

We believe it to be incumbent upon any modern theory of hippocampal function that it should present a picture that is consistent with the position of the hippocampal formation within each of the very many distinct neural systems of which it is a component. The temptation, of course, is to ascribe to the hippocampus a different function with respect to each of them. However, the internal anatomy of the hippocampus (Fig. 1.3) suggests a high degree of uniformity in its processing, while its external connectivity (Appendix 4) suggests a high degree of topographic mapping. In constructing our own theory, therefore, we have searched for fundamental cybernetic principles of operation of the hippocampus which are the same for its relations with all its afferent and efferent connections. We have also attempted to derive, for each of the several levels of the hippocampal formation, a single principle of operation which is consistent with the principles ascribed to the other levels. The apparently different psychological functions of the septo-hippocampal system in different situations should then derive mostly from the particular structures with which it interacts at different times.

Principle 6. Any theory of hippocampal function should account for all the different effects of septo-hippocampal disruption and all of the observed correlations of cellular activity in terms of a single consistent set of principles. It should also relate these to the known functions of the structures immediately afferent and efferent to the septo-hippocampal system.

9.6 THE ANATOMY OF THE SEPTO-HIPPOCAMPAL SYSTEM

If we are to relate hippocampal function to its afferents and efferents, we must first highlight some key aspects of the anatomy of the septo-hippocampal system and of the structures most closely related to it. We can trace our way (antero- or retrogradely) ever deeper into the brain from several different starting points via a large number of independent routes all leading to the hippocampus: dorsal periaqueductal grey–hypothalamus–amygdala–hippocampus (Chapter 6, Appendix 2); primary sensory systems–association cortex–parahippocampal cortex–entorhinal cortex–hippocampus (Appendix 4); nucleus reticularis pontis oralis–supramammillary nucleus–medial septum–hippocampus (Appendix 5); hypothalamus–cingulate cortex–hippocampus (Chapter 6, Appendix 3); primary motor cortex–supplementary motor cortex–prefrontal cortex–hippocampus (Chapter 6, Appendix 3); cerebellum–retrosplenial cortex–hippocampus (Appendices 3 and 5).

There are two general points to note about the detailed anatomy (Fig. 9.4). First, the septo-hippocampal system has major connections with many motor programming areas, including extensive subcortical connections. The mapping of these areas into the different levels and parts of the hippocampus (e.g. Risold and Swanson 1996; Fig. 9.5) provides the only clear source of topographic organization within the hippocampal formation. Relatively speaking, then, hippocampal links with pure sensory processing, and potentially with such

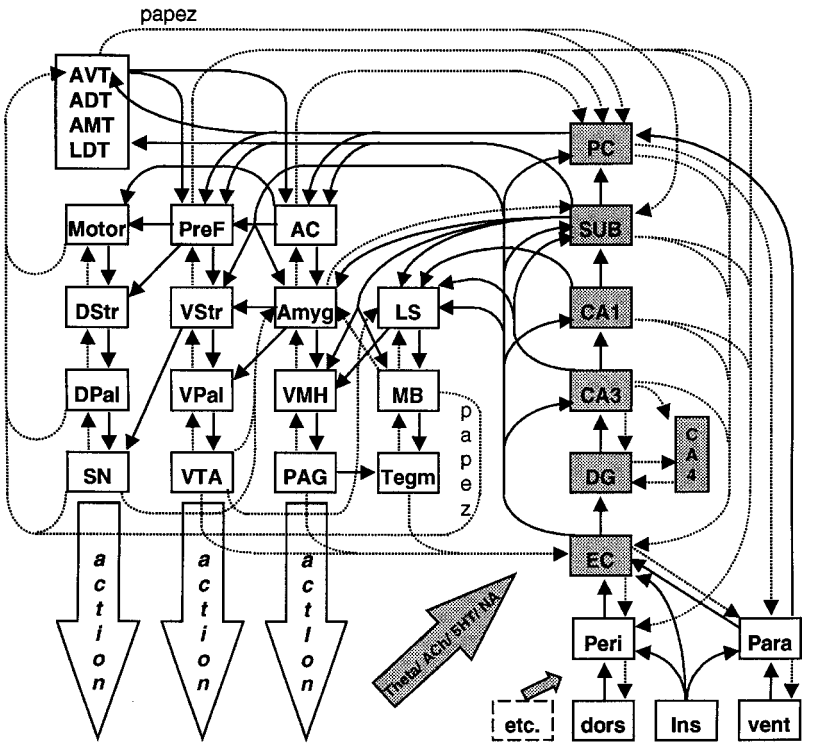


Fig. 9.4 The general organization of the hippocampal formation and its connections (see also Fig. 10.6). The hippocampal formation can be viewed as an essentially linear, unidirectional set of modules (EC, DG, CA3, CA1, SUB, PC) embedded in systems which have largely bidirectional connections. It receives (shaded large arrow), as a whole (shaded boxes), phasic theta-controlling GABAergic input together with cholinergic, serotonergic, and noradrenergic inputs which gate or modulate the occurrence of theta. (DG, CA3, CA1, SUB, and PC can also be viewed as receiving parallel input from the subcortical theta controlling systems and from EC.) CA3, CA1, SUB, and PC all provide feedback to EC. The hippocampal formation is also embedded in a range of external loops of which the Papez circuit (papez) is the most famous example and can be viewed here as potentially involving EC, DG, CA3, CA1, SUB, LS, MB, AVT, PC, thence back to EC. We view the hippocampal formation as interacting with a range of modules (oriented vertically in the diagram), each with units at a number of cortical and subcortical levels (shown by the rows in the diagram, where lower rows are more caudal in the brain than the higher rows). We see these modules as principally concerned with the organization of goal-directed action. (Systems such as the cerebellar circuits involved in the control of eye-blink conditioning are omitted here for simplicity, but can be viewed as having an essentially similar organization.) EC receives input from the lower stages of most of these systems and also receives input, along with PC, from the higher stages. EC also receives input, usually relayed by Peri and Para, from the higher levels of the dorsal and ventral streams of sensory cortex. AVT, anteroventral thalamus; ADT, anterodorsal thalamus; AMT, anteromedial thalamus; LDT, laterodorsal thalamus; Motor, motor and premotor cortex; DStr, dorsal striatum; DPal, dorsal pallidum; SN, substantia nigra; PreF, prefrontal cortex; VStr, ventral striatum; VPal, ventral pallidum; VTA, ventral tegmental area; AC, anterior cingulate cortex; Amyg, amygdala; VMH, ventromedial hypothalamus; PAG, periaqueductal grey; LS, lateral septum; MB, mammillary bodies; Tegm, dorsal and ventral tegmental nuclei; PC, posterior cingulate; SUB, subiculum; CA1, CA3, CA4, subfields of Ammon's Horn (Cornu Ammonis); DG, dentate gyrus; EC, entorhinal cortex; Peri, perirhinal cortex; Para, parahippocampal cortex; dors, vent, dorsal and ventral streams of sensory cortex; Ins, Insula.

memory as can be dissociated from action, are modest, although not negligible. Second, the septo-hippocampal system sends output (often from the subiculum) to more sensory levels of many of these systems (or to polymodal association cortex) and receives input (often to the entorhinal cortex) not only from those sensory levels but also from the more motor levels of many of those same systems. Together with the fundamentally linear flow of information within the septo-hippocampal system itself, this leaves us with a picture of the hippocampus as a device which circulates information recursively through its target structures. This is reminiscent of the prefrontal circuitry involved in the 'rehearsal' which constitutes working memory (see above and Appendix 3). But we believe it has essentially opposite functions: prefrontal recursion maintaining neural activity and hippocampal recursion suppressing it (albeit selectively).

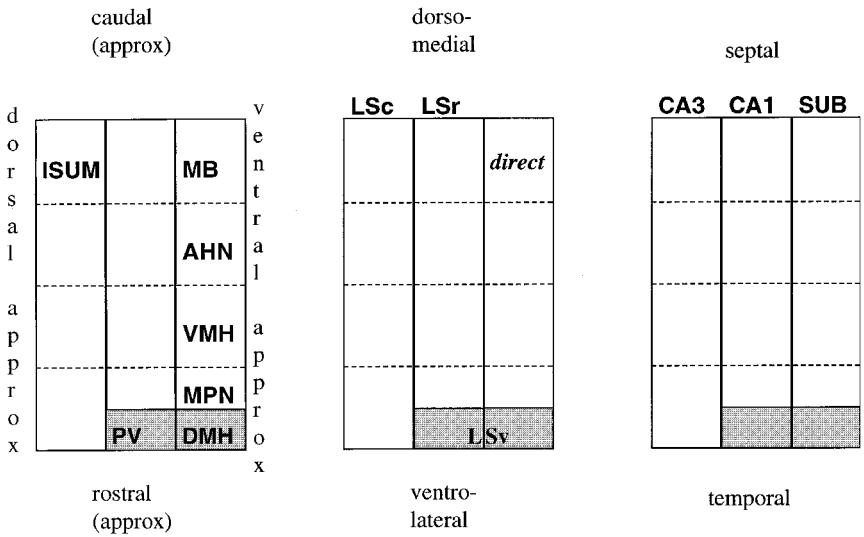


Fig. 9.5 Topographic mapping of hippocampal formation into the septal area and hypothalamus. The hippocampal formation can be viewed as a two-dimensional sheet (right panel) with unfolding of the cell layers and removal of the curvature of the hippocampus resulting in the top of the sheet being the septal pole of the hippocampus and the bottom being the temporal pole, the left-hand side of the sheet being CA3 and the right-hand side being SUB. (The central portion is drawn as area CA1, but recent data, including evidence that the amygdala connects to an area which is essentially the CA1–SUB border, suggest that it may be better to view the hippocampus as made up of a larger number of strips than the three shown.) The hippocampus is topographically mapped into the lateral septal area which can also be viewed as a two-dimensional sheet which, with the proper bending and orientation (middle panel), provides a direct correspondence with the hippocampal sheet (e.g. the shaded temporal portion of CA1 and SUB projects to LSv). The floor of the brain in the region of the hypothalamus can be viewed similarly (left panel), with the sheet essentially standing on edge (lSUM is largely dorsal to MB), with septal portions of the hippocampus projecting via the septum to caudal hypothalamic areas and temporal portions to rostral (giving approximately similar path lengths in the two cases). The most septal quarter of SUB appears to project directly to MB without relaying in the septum. Abbreviations as for Fig. 9.4. c, r, v, caudal, rostral, ventral; AHN, anterior hypothalamic nucleus; MPN, nucleus medial preoptic nucleus; DMH, dorsomedial hypothalamic; PV, periventricular nucleus; lSUM, lateral supramammillary nucleus. (Redrawn following Risold and Swanson 1996.)

Principle 7. A theory of hippocampal function must account for the fact that major connections of the hippocampus (like the hippocampus itself) are phylogenetically old and often provide links to systems directly concerned with action, many of which are subcortical. Further, the anatomy of the hippocampus and its connections is such that information will tend to circulate in long loops, with the hippocampal formation as a key element forcing the information to flow in only one direction (dentate–CA3–CA1–subiculum).

A final point, which leads to the greatest uncertainty in the current version of our theory, is that the hippocampal formation has many levels. A complete theory should explain the functions carried out by each of the entorhinal cortex, dentate gyrus, CA3, CA1, subiculum, and posterior cingulate cortex. The different levels are unlikely to exist simply so as to pass unaltered information on to one another; each must allow some particular transformation or integration of its inputs. We propose that the hippocampus is a system of logical gates which allows different types of information to progress to different points of the circuit and hence to produce (or in many cases not produce) outputs from different levels of the system subject to different conditions. We are fairly certain that the hippocampus constitutes a set of such logical gates. On the other hand, we are far from certain that our specific assignments of input values to, transformations produced by, and outputs from the different levels will prove accurate when there are sufficient new data to test our ideas directly. Our detailed specification of functions for particular components (Chapter 10) is made quite consciously, therefore, to encourage experiments able to disprove the specific details, and to make clear the nature of the theory itself. While the mechanisms we postulate are speculative, it should be noted that other current theories (with the sterling exception of the O'Keefe and Nadel spatial mapping model) do not assign specific information processing functions to the different parts of the hippocampal formation at all.

Principle 8. Each separate level of the hippocampal formation represents a separate logical gate. A distinct function must be assigned to each of these gates and must account for the outflow of the resultant information from areas CA3 and subiculum, in particular.

Before leaving the anatomy we should remind the reader of the power of even simple recursion in parallel networks, and the possible role of multiple levels in such recursion. This was discussed extensively in Section 1.9, with Marr and Poggio's global stereopsis algorithm as an exemplar (Fig. 1.6, p. 26). This leads to:

Principle 9. Recursive external and internal connections and multiple levels in the hippocampus allow for functionally sophisticated output as a result of iterations based on very simple computations.

It should also be noted that the hippocampus is provided with quite complex intrinsic interneuronal circuitry (Fig. 9.6) that may be necessary to maintain stability in what, otherwise, could be essentially unlimited positive feedback loops.

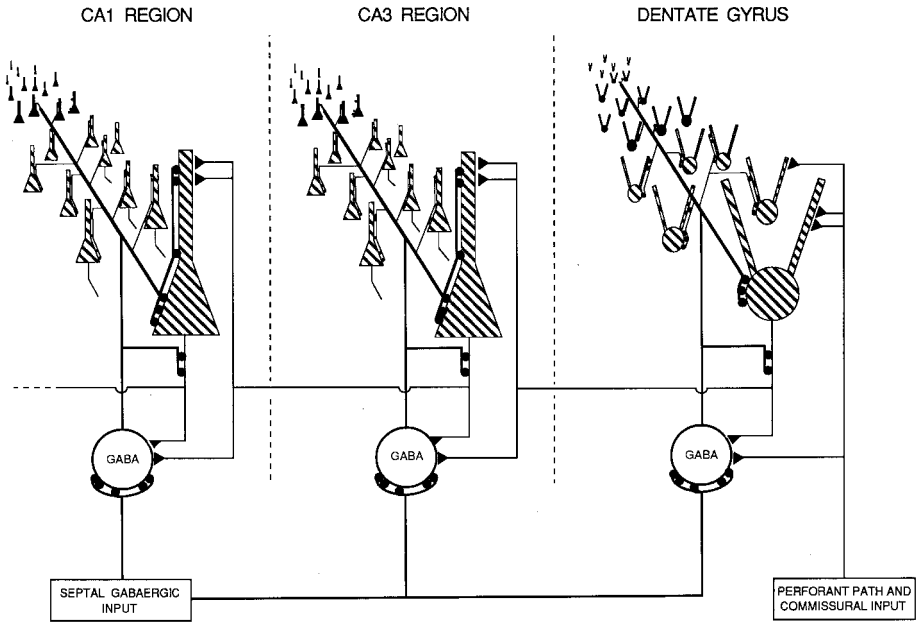


Fig. 9.6 Connections of hippocampal interneurons. Cross-hatched neurons, principal cells. (From Freund and Antal 1988.)

9.7 LONG-TERM MEMORY

It is generally agreed that, for very long-term memories (and, in our view, for all durations of memory), the relevant associations are located outside the hippocampal formation. The role of the hippocampal system in memory is then usually seen as involving interaction with sensory, particularly polymodal, cortex. We agree that this will often be the case. But, if one equates the formation of memory with learning, subcortical areas such as the amygdala and nucleus interpositus should also be included with neocortical areas as sites of memory formation. Whatever distinctions might be made at the psychological level, the plasticity seen in the amygdala during avoidance conditioning or in the nucleus interpositus during eye-blink conditioning does not seem to differ in any fundamental way from cortical plasticity. In many respects, then, we can see memory as being distributed throughout the brain.

Given this view of simple associative memory, what could be the function of the hippocampus? We rule out working memory, both because of the data reviewed in Chapter 8 and because (following Goldman-Rakic) we have attributed this function (Chapter 6; Appendix 3) to the prefrontal cortex. We rule out long-term memory of all types, as this is intact in H.M. and related amnesic cases. We rule out intermediate memory, because (Chapter 8) of the presence of intrusion errors in amnesics' responses and the effects of interference.

We propose the following scenario. The associative rule instantiated in long-term potentiation requires that a stimulus representation of a goal be activated if any strength-

ening is to occur. Where only one such stimulus representation is available for conditioning, then only the connections of that stimulus can be strengthened. If, therefore, the task (e.g. in an eye-blink conditioning paradigm) is set up so that there can be no possible strengthening also of inappropriate connections (or so little possibility that no inappropriate response is generated able to compete with the one required), then memory formation can occur, without hippocampal involvement, in the region where the relevant sensory–motor connections are located. However, where the task demands become sufficiently complicated, the distributed activity in the extra-hippocampal net will represent a number of incompatible concurrent goals. Hence, a number of different and functionally conflicting pathways could have their connections strengthened. It is the business of the hippocampal system to resolve this conflict between goals and to prevent too wholesale a strengthening of associative links.

Principle 10. The hippocampal formation is at no time a memory store, intermediate or otherwise. Nor does it aid in the formation of specific strengthened connections. The primary function of the hippocampus is inhibition. In the context of memory experiments this inhibition represents the suppression of incorrect alternatives and is necessary in some variants of some paradigms if consolidation is to result in the correct recall of only a single alternative. In effect, output from the hippocampus enhances the negative associations of inappropriate stimuli.

A possible model for the operation of the hippocampus here is suggested by the models of global stereopsis and language processing discussed in Chapter 1 in relation to the recursive anatomy of the septo-hippocampal system. In these, input is clamped onto a recursive network. Iteration of the information within the net allows a figure to be extracted from a ground, or a correct item to be selected from a pandemonium, in a process which progressively restricts the output from the network. The recursive links of the hippocampus with all its closest neighbours would, with suitable connections, allow it to 'sharpen up' the encoding of stimulus information, and hence to resolve the conflict between the alternative possibilities.

Thus, where a set of neural links A–B–C is required in contrast to A–C–B, the hippocampus would act, during the initial formation of the A–B and B–C links, to prevent the (extremely likely) formation of the highly undesirable A–C and C–B links. The A–C and C–B links are likely because simple associative rules, like those embodied in long-term potentiation, will operate whenever items A, B, and C are all concurrently present (in the environment or in active memory), as they must be if a relation between them is to be perceived. Thus the hippocampus acts asymmetrically. It interacts with the formation of all current possible relations to suppress those which have more adverse associations. That is, it enhances the value of the negative reinforcement which will accompany activation of the A–C or C–B links. Even in memory, then, the hippocampal formation acts to increase negative bias and, consequentially, produce inhibition. A process of this kind has some of the flavour of lateral inhibition, but the recursive mechanism is inherently much more powerful and distinctive in producing non-linear results. Hippocampal recursion need not result in a simple 'winner-take-all' output, in which the initially most active link would be enhanced and others suppressed. The hippocampal

formation will act to increase only affectively negative associations of a particular goal, and hence to inhibit that goal in a fashion unrelated to its prior level of activation. Furthermore, the effects of the hippocampus on any particular cycle of a recursive sequence will depend on the results of the immediately preceding cycle, which will change the weights of the inputs fed back to the hippocampal input units. Thus, the end-result depends on the pattern of weights of negative associations enhanced by the hippocampus across a range of different efferent targets and would not depend on the initial level of activity in any target.

Active recursion is likely to have an additional advantage over simple lateral inhibition. The extent of the resultant inhibition can be adjusted from occasion to occasion in relation to the requirements of the task. This means that its effects will be independent of the intensity of the individual inputs to the hippocampal formation, and that they can be gated out if this is required (hence, we believe, the requirement for long-term potentiation in the hippocampus).

We have just suggested that the hippocampal formation is a device which can resolve conflicting memory representations. But our earlier equation of perception with action implies that a precisely similar hippocampal operation in relation to neural networks which code for movement would have the net effect of behavioural inhibition. Although we did not detail the machinery earlier, the mechanism suggested for memory processing by the hippocampus is sufficient to support the functions already proposed for its interactions with both the defence and eye-blink conditioning systems.

The connections of the hippocampal formation with each of amygdala, cingulate cortex (and hence nucleus interpositus), prefrontal cortex, and rhinal cortex can then be viewed as operating in the same basic way in which it interacts with polymodal association cortex. The hippocampus would circulate information between itself and those areas so as to accentuate peaks of activation in a way roughly equivalent to lateral inhibition in sensory systems, but with a power and sophistication equivalent to that of the global stereopsis network or the linguistic network discussed in Chapter 1 (with their 'clean up' units able to extract figure from ground), and a non-linearity dependent on the fact that it acts mainly to increase existing negative biases and does not interact directly with current positive biases.

Principle 11. The longer-term effects of hippocampal function are achieved through the recursive circulation of information between it and its efferent targets. This processing is essentially identical, whether the target is an area which processes predominantly sensory or predominantly motor information. The recursive operation of the hippocampal formation is akin to the extraction of a figure from a ground and depends on amplification of affectively negative associations.

9.8 ROLE OF THE SEPTO-HIPPOCAMPAL SYSTEM IN SENSORY PROCESSING

As one follows a sensory system from the periphery, deeper into the brain, the information encoded is initially unimodal, but then, for example in parietal cortex, this gives way to

multimodal encoding. If we follow, say, visual information from V1 to V2, V4, inferotemporal cortex, perirhinal and parahippocampal cortex, entorhinal cortex and, finally, hippocampus, we could, by extrapolation, conclude that the hippocampus is concerned with producing the very highest order abstractions from multimodal sensory information. However, the bulk of the evidence is as much against a purely perceptual as against a memorial or motor function for the hippocampus. Furthermore, the bulk of the single cell data suggests that many hippocampal cells are non-specifically multimodal. Rather than firing to only one stimulus which has a specific combination of, say, visual, auditory, and olfactory properties, the cell fires to any of several stimuli whether visual, auditory, or olfactory (Appendix 6). Furthermore, anatomically and developmentally, the hippocampus appears distinct from the higher levels of association cortex.

Principle 12. The hippocampus is best viewed as a device which modulates activity in sensory systems, but is not itself a higher level of the classical sensory systems.

An important route by which this modulation of sensory processing can take place is via the nucleus accumbens (to which the hippocampal system projects from the subiculum and entorhinal cortex), and thence on successively (via a series of GABAergic pathways) to the ventral pallidum, nucleus reticularis thalami, and the whole array of thalamocortical sensory processing loops. Analysis of the likely mode of operation of this circuit suggests that activation of the subiculo-accumbens projection can boost thalamocortical sensory processing in all modalities, while analysis of the likely behavioural effects of such activation suggests that it is particularly novel stimuli and stimuli with strong associative significance that are picked out for enhanced sensory processing in this way (Gray *et al.* 1997). It is possible that this route is responsible for conscious processing of anxiety-related stimuli (Gray 1995), but we shall not pursue that possibility in this book.

9.9 THE MISMATCH DETECTION SYSTEM

Conflict between goals, in the simplest sense, will occur when two incompatible goals are present concurrently. We have presented a model in which the hippocampus (or strictly its interaction with its target structures) can detect, as well as resolve, such conflict. There is also reason to believe that the hippocampus detects the more subtle conflict which occurs when there is a mismatch between observed and expected events.

The idea of detection of conflict is inherent in Vinogradova's (1975) view that the septo-hippocampal system mediates the orienting response and its habituation (Appendix 6). This view is close to the more general one proposed by Douglas (1967) that the hippocampus gates out redundant stimuli from the control of behaviour. Both these views are in many respects complementary to the view that the hippocampus is involved in motor components of behavioural inhibition (Kimble 1969).

Vinogradova's view gains support from the association of low-frequency theta with responses to novelty and her (Vinogradova 1975) observations of hippocampal and lateral septal unit responses to novel and familiar stimuli. These show that the septo-hippocampal system receives the information it would require if it were indeed to organize orienting

behaviour or eliminate responses to unimportant stimuli. On the other hand, only a few of the findings in lesioned animals support this specific position, for example the loss of Kamin's blocking effect (Solomon 1977; Rickert *et al.* 1978) or of latent inhibition (Ackil *et al.* 1969; Weiss *et al.* 1974; Solomon and Moore 1975; see Buhusi *et al.* 1998 for review). Indeed, a critical point against the most substantive version of Vinogradova's hypothesis is that there is very little evidence of change in simple orienting responses or their habituation after damage to the septo-hippocampal system (Appendix 8). Most of the relevant data can be more easily accommodated by the generalization that motor responses to novel stimuli (as to stimuli of other kinds) are released from inhibition in the lesioned animal.

It seems likely, then, that the apparently simple neuronal habituation described by Vinogradova (1975) underlies something more complex than simple habituation at the behavioural level. This inference is strengthened by a consideration of the experiments on Kamin's blocking effect. The normal animal fails to develop an association to a second CS if this is added to a first one which already predicts the same UCS (Mackintosh 1974), but animals with lesions of the hippocampal formation respond to the added CS (Solomon 1977; Rickert *et al.* 1978; Buhusi *et al.* 1998). This is probably not due to a simple delay in the habituation of the response to the first CS, since Rickert *et al.* (1978) failed to alter the effect of hippocampal lesions by doubling the number of trials of training on the first CS. It is possible that, in the normal animal, blocking depends on a process that one might call 'instructed habituation'; that is, as a result of the discovery on the first compound-stimulus trial that reinforcement conditions have not changed, the redundant stimulus is identified as one to which attention need not be paid (Mackintosh 1978). Thus, hippocampal lesions may disrupt this more complex type of habituation. This hypothesis is consistent with Douglas's (1967) description of the hippocampal formation as a device for gating out stimuli that are not predictive of reinforcement.

Let us consider precisely how, according to Vinogradova, the hippocampus could perform this function. As we summarize her position in Appendix 6, the hippocampal formation receives sensory information by two independent routes. Stimuli encoded by the reticular formation are relayed by the medial septum to the hippocampus (where they represent a highly general alerting signal which has lost much of the specific qualities of the initiating stimulus). The same stimuli are processed by the cortex and, if a model of the stimuli is already present in the cortex, the actual occurrence of the predicted stimulus is also sent via the entorhinal cortex to the hippocampus, where, according to Vinogradova, long-term potentiation of the perforant path input to the dentate gyrus prevents further processing by the latter structure. On this view, the hippocampus acts as a comparator of observed with expected input. If there is input via the medial septum but not entorhinal cortex, the stimulus can be classified as novel. Equally, if there is input from entorhinal cortex but not from the septum, an expected stimulus can be classified as absent. Given either type of mismatch, a response is required from the hippocampus. Note that while the input from the medial septum here permits the hippocampus to detect novelty (and hence to detect potential conflict with ongoing motor programs), we need not regard it as a signal of novelty *per se*. Rather, in keeping with the reticular origin of the information, it should be viewed merely as a signal that some stimulus has occurred and, as we emphasize in Appendix 6, that some response may be required. The entorhinal signal can then be viewed as indicating that

an appropriate response will be forthcoming, and the devaluing of a redundant stimulus would be equivalent to the suppression of the competing responses to which it might otherwise give rise.

Although this is not emphasized by Vinogradova, the septal signal must be of a fairly crude type, as the medial septal area has very few neurons with which to encode fine distinctions and, moreover, sends its output fairly diffusely to all subregions of the hippocampal formation, including the entorhinal and posterior cingulate cortices. Indeed, given the known properties of the reticular system and the reaction of the theta system to stimulation of the midbrain, it is possible that the septal signal carries no specific information at all. Rather, at all times, the input from the reticular formation via the septum can be seen as providing an indication of the general level of activation summed across a number of subcortical systems. However, given the topographic organization of the septal input to the hippocampus, the topographic mapping of the hippocampus into hypothalamic 'goal space' (Fig. 9.5), and the detailed variations in the firing patterns of individual septal neurons, it seems likely that some aspect of goal type or at least valence is encoded as well. In a novel situation, therefore, the hippocampus receives information that something is going on (and in essence how important that something is), but has no information as to what. This discrepancy, if unchecked, results in the activation of behaviour designed to make good the information deficiency. Operation of the output from the subiculum via nucleus accumbens to thalamocortical sensory processing systems, as indicated in the previous section, may in part underlie this reaction pattern.

The Douglas–Vinogradova 'attentional' view of the generation of behavioural inhibition by the hippocampus is closely related to two other hypotheses.

The first is the hypothesis that the septo-hippocampal system mediates responses to non-reward (Gray 1970a). Since non-reward involves the omission of an expected stimulus (the anticipated reward), this hypothesis is probably not fully dissociable logically from Vinogradova's; empirically, Amsel (1972) has in any event made out a strong case for the view that novelty and non-reward elicit essentially similar responses. Since these responses include inhibition of ongoing behaviour and increased attention to the environment, the non-reward hypothesis is equally close to the motor and attentional versions of a general behavioural inhibition theory.

The second close relative of the Douglas–Vinogradova position is Gaffan's (1972) recognition memory hypothesis of hippocampal function. Since this proposes that the hippocampus is a mechanism for detecting novelty and/or familiarity, it is at first sight indistinguishable from Vinogradova's orienting reflex hypothesis. It can be distinguished, however, from Douglas's (1967) position because, according to Gaffan, the hippocampus is involved in the recognition of familiarity, whether this results in the production or the inhibition of responses, whereas Douglas supposes it to be concerned only with the latter. The data (Chapter 8; Appendix 8) give greater support to Douglas: hippocampal memory deficits consist largely of false positives rather than false negatives, and these are heavily influenced by sources of interference in the experimental task. In particular, trial-unique delayed object discriminations are intact after hippocampal lesions despite their requirement for a relative familiarity discrimination.

Gaffan's concept of recognition memory is very close to Olton's working memory hypothesis (Chapter 8). While we reject both, there is an important feature which they

share, which sets them apart from Vinogradova's simpler treatment of novelty, and with which we agree: that is, their concern with renewable lists. These consist of a sequence of items that recur, require a response when they first occur, then for a period of time or during certain experimental circumstances do not require a response, and then again require a response on some later occasion. (An example is the radial-arm maze, in which during one trial the animal needs to visit each arm in turn, without re-entry into any given arm, and is then required to repeat this process on a later trial.) In Gaffan's case, renewable lists are a feature of experimental practice, rather than, as in Olton's, central to the theory; but it is the experimental results which we must in any case accommodate. Renewable lists require the subject to answer not just the question 'is this stimulus novel or familiar?' but also 'is this stimulus more or less familiar than that one?' or 'is it *relatively* novel (within the last day, hour, or few minutes)?' Thus, if the hippocampus is the organ responsible for successful performance in tasks of this kind, it must be given additional properties besides those required for its higher-order control of habituation. We must ask, therefore, what it is about lists, and especially renewable lists that require recognition memory as defined by Gaffan (1977a,b), which provokes a particularly profound and enduring deficit in the performance of hippocampal animals (Olton *et al.* 1979a,b)?

A natural link exists between Gaffan's recognition memory hypothesis and earlier research on the orienting reflex (Sokolov 1960). This research showed that, although the response to a repeated neutral stimulus habituates, habituation can be prevented for an indefinite period by turning the stimulus into a Pavlovian CS. Note also that this treatment guarantees that the stimulus continues to evoke unit responses in the hippocampal formation (Appendix 6). Observations such as these imply that the habituation of the orienting reflex (both behavioural and 'hippocampal') is under higher controls that are sensitive to the associative significance of stimuli. These controls might be of cortical origin (Sokolov 1960) or they might form part of the hippocampal circuitry itself; and there are of course other possibilities.

Performance on a renewable list, of the kind used by Gaffan or Olton, might depend, then, on a type of higher nervous control over a hippocampal circuit whose basic function is that of mediating the production or habituation of orienting responses under simpler conditions. For this kind of control to work, it would be necessary that habituation of the orienting reflex could be both cancelled (at each new presentation of a previously encountered list) and expedited (when an item within a given list presentation has been sufficiently processed for the current trial). It is possible that the second of these processes is also responsible for 'instructed habituation' in the Kamin blocking effect, as discussed above. By analogy, therefore, we can call the first process (cancellation of habituation) 'instructed dishabituation'.

In all of the above, matching of septal input with a 'model' coming from entorhinal cortex can be seen as preventing the behavioural inhibition and orienting/information gathering which would otherwise occur. This implies

Principle 13. Matching input from the medial septal area and entorhinal cortex will normally result in output which, in effect, constitutes a 'familiar-ignore' signal. Mismatch (either septal input in the absence of entorhinal, or entorhinal in the absence of septal) will result in information gathering behaviour to resolve the discrepancy.

Note that familiarity, *qua* familiarity, is encoded elsewhere. Also note that this familiar-ignore signal, while it is fundamental for at least some aspects of hippocampal processing, represents an *input* to the hippocampus proper. Hence it cannot reflect the major purpose of the circuitry of this structure. The entorhinal cortex, then, allows for instructed habituation, latent inhibition and for at least some aspects of the partial reinforcement extinction effect (which can be viewed as a special case of latent inhibition to the stimuli of reward omission; see Appendix 9 for a discussion of the relevant data). However, a system capable of forcing response programming mechanisms to ignore all familiar stimuli could itself be a liability; it will require suppression when the stimuli are, in fact, important. We turn to this problem in the following section.

9.10 AMINERGIC GATING SYSTEMS

Even when viewed as a means of reducing interference between conflicting goals rather than as a full-blown memory system, our description of the hippocampal formation has provided a highly cognitive, memory-oriented view of the hippocampal formation. However, 'goals' and 'plans' are not affectively neutral. It is time, therefore, to reaffirm the links of the hippocampus with emotion, and to consider the contribution of the diffuse ascending subcortical systems to its operations.

The cholinergic, serotonergic, and noradrenergic inputs to the hippocampus originate in small nuclei which send collaterals to much of the forebrain. In our account of these systems we have tried to identify the common, fairly simple, signal which each sends to all of its target structures, and to account for the different uses to which this common information is put by the different targets (Appendix 10). In particular, we noted that the serotonergic input (whether interpreted as one of conditioned punishment or, better, of motor readiness) to the dorsal periaqueductal grey and amygdala has opposite effects on these two structures, thus providing a source of coordination of hierarchical elements of the system (Chapter 6). We have argued that all the aminergic systems are concerned, but in somewhat different ways, with the focusing of 'attention' or, better, with reduction in signal-to-noise ratios within the septo-hippocampal system. We also argued for a similar modulatory function, but in the temporal domain, for the ascending theta system and particularly its GABAergic component.

Principle 14. The three aminergic systems appear to provide largely undifferentiated signals (perhaps translatable in psychological terms as 'arousal' or 'attention' of different types) which increase the signal-to-noise ratio between different afferents both within the hippocampus and within the defence system as a whole.

Where the simple associative significance of stimuli is the crucial feature, the noradrenergic input to the hippocampus appears to be important (Appendix 10). Thus, this sends to the hippocampus a signal that indicates that a reward, punishment, or novel stimulus is present. In all these cases, the stimulus must be treated as important and must, potentially, be capable of interrupting ongoing motor programs. This does not necessarily, however, entail output from the hippocampus. In the case of reward, we

note in Appendix 10 that input from the locus coeruleus could act, in effect, to prevent input from the entorhinal cortex from affecting subsequent elements of the hippocampal trisynaptic circuit. Thus it is only an unexpected reward or omission of an expected reward which will affect later stages of hippocampal processing. The same requirement for unexpectedness will be true, of course, of the capacity of a novel neutral stimulus to produce hippocampal output.

Punishing stimuli, however, require rather different treatment. The noradrenergic input to the hippocampus appears (Appendix 10) functionally more concerned with reward and novelty, while the serotonergic input appears more concerned with punishment, despite the fact that their final effects on hippocampal electrophysiology seem fairly similar and neither is tied to a particular affective valence (see below). Here, we should note an asymmetry. It is possible to present a CS for punishment (CS-Pun+; see Chapter 3 for terminology) in the absence of prior conditioning; however, a CS for reward omission (CS-Rew–) can be presented only against a prior conditioning history of reward. The hippocampus, therefore, needs to be able to distinguish between presentation of a reward and presentation of a punisher. While it could do so via cortical inputs, it might do so also by comparison of noradrenergic and serotonergic inputs. Thus, both punishment and reward activate the former but, since punishment but not reward activates the latter, the ambiguity can be resolved.

This brings us to the issue of the distinction between CS-Pun+ and CS-Pun– (Chapter 3). The serotonergic input to the hippocampus is from collaterals of neurons which innervate, among other structures, the periaqueductal grey and amygdala (Fig. 6.6). The information it carries, therefore, must be fairly non-specific. In the case of the periaqueductal grey, release of serotonin blocks functional output, whereas in the amygdala and hippocampus such release increases it. This differentiated pattern of change is consistent with the requirement to inhibit undirected escape when directed escape or avoidance is possible (Chapter 6).

What, however, of conjoint serotonergic input to the hippocampus and the amygdala? Whether this input is essentially inhibitory, excitatory, or increases the signal-to-noise ratio, it is only of concern to the hippocampus if it is associated with a CS-Pun+ and not a CS-Pun–. Here we can postulate a function for the link between the amygdala and the hippocampus. Serotonergic input, as suggested by Graeff (1994), signals the anticipation of an aversive event (CS-Pun+ or CS-Pun–) to both hippocampus and amygdala (or, better, it signals the readiness to respond to whatever event is currently predicted, while the amygdala codes the aversive nature of the event; see Appendix 10). In the presence of a CS-Pun–, the amygdala will be activated to produce an avoidance response and can signal this to the hippocampus. From the point of view of the hippocampus, then, CS-Pun+ could, in principle, be signalled by input from the raphe which lacks concurrent input from the amgydala. However, a more likely scenario is that CS-Pun+ will be signalled by concurrent activation of the amygdala and some other area which is activated to produce an incompatible approach response. The hippocampus can then act to inhibit this approach response.

The rather glib phrases 'signals of reward, punishment, and novelty' and 'signals of punishment' which we have applied to the noradrenergic and serotonergic systems, respectively, need closer inspection. First, we should note the non-specificity implied. In the noradrenergic case, we presumed that the hippocampus (or any other target structure)

could not differentiate between the three different relevant types of important event on the basis of this input alone. Even in the serotonergic case, the small number of neurons in the raphe suggests that there will be little differentiation between different classes of negative reinforcer; and we argue in Appendix 10 that it is motor readiness in general, not aversion, which is being coded. We are clearly, therefore, dealing with modulatory systems rather than ones which carry specific information.

We have also been talking as if the ascending inputs to the septo-hippocampal system (and the amygdala, which also receives both noradrenergic and serotonergic input) were simple unidirectional inputs. However, the key architectural feature of the septo-hippocampal system and structures closely related to it is recursiveness, due (as we have noted) not only to the bidirectional connections between individual structures, but also to longer loops which apparently return information to structures from which it originated. We have already discussed the general properties of such recursive architectures, and will discuss shortly the specific properties of this architecture in the context of the septo-hippocampal system. For the present we should note that, while the monoaminergic inputs may carry signals resulting from reward, punishment, and novelty, they should not be perceived as signalling reward, punishment, or novelty as such. Rather, their capacity to increase signal-to-noise ratios should be seen as setting parameter values at various points in the recursive loops of the overall hippocampal system, so determining what information these loops preferentially process. In this way, for example, the monoaminergic afferents might allow other, more specific inputs related to reward, punishment, etc. to have a greater control over behaviour than they would otherwise be able to exert.

In Chapter 6, we pointed out that lesions of the amygdala impair active as well as passive avoidance, while hippocampal lesions impair only the latter. This is best accounted for by the notion that the amygdala is not only the source of action in the form of avoidance, but also the source of perception of threat. It can be viewed, then, as passing to the hippocampus highly specific information which the hippocampus then uses for inhibitory purposes. Equally, the hippocampus could be seen as passing back to the amygdala cleaned-up versions of the same information which would have the net effect of increasing avoidance responses in cases where conflict is present. Thus, the specific signals of punishment (and probably, in some cases, of reward) can come to the hippocampus from the amygdala, while activity in the noradrenergic and serotonergic systems reflects increases in more general aspects of 'sensory attention' and 'motor attention', respectively, without which hippocampal processing cannot operate efficiently.

Principle 15. The aminergic inputs to the septo-hippocampal system reflect relatively undifferentiated 'attentional' signals which alter the signal-to-noise ratio in the system. Each signals a different type of event or requirement.

9.11 THETA ACTIVITY

We might have been thought to cover the role of the theta control system in the previous section, when we discussed septal input to the hippocampal formation. We

there attributed a function to the septal input itself (or at least its cholinergic aspects). However, we gave no account of the occurrence of theta activity, or of the fact that the frequency of theta appears to be determined simply by the total sum of inputs to the supramammillary nucleus and the dorsomedial hypothalamic nucleus (and/or any other nuclei linked to them). Furthermore, the control of frequency appears to be through small numbers of local inhibitory interneurons which regularly clamp the output of much larger numbers of projection neurons (Appendix 5). Given this mode of organization, changes in the control of theta would be expected to modify the temporal packaging of information in the ascending theta system, but will not necessarily change any topographic encoding of that information. Similarly, alteration of the frequency or presence of theta in the hippocampus and entorhinal cortex would be presumed to leave intact the (much richer) topographic encoding of information in these areas (these issues are discussed at greater length in Chapter 10, Section 10.5). Of course, where loss of theta has resulted from medial septal lesions, and especially electrolytic medial septal lesions, there would be additional loss of the specific information transferred to the hippocampus via the septum, on top of the loss of temporal packaging of all hippocampal–entorhinal information by theta. However, even under these conditions many of the important connections of the septo-hippocampal system would remain intact.

Thus, neither medial septal lesions nor injections of anxiolytic drugs would be expected to have effects identical to those of hippocampal lesions. In particular, theta is not seen as essential to all hippocampal function. We have mentioned the possibility that the hippocampus carries out its functions by recursive recycling of information between itself and target structures. If this is the case, phase-locking of cellular activity (which would be evidenced in many cases by the presence of theta rhythm) could be expected to enhance the cleaning-up processes being carried out by the recursion, clearing the results of a previous iteration before starting on the next. However, the absence of such phase-locking would not necessarily be catastrophic; it could merely reduce the efficiency of processing, not eliminate it altogether.

Principle 16. Theta activity in the hippocampal formation is the result of phasic drive from subcortical systems, gated by tonic cholinergic and serotonergic inputs. The phasic aspect of theta is likely to be important for hippocampal function only when optimal performance of the system is required.

9.12 CONCLUSION

We have now enunciated a series of principles which we believe are solidly based on the available data and against which, therefore, all theories should be judged. While these place severe limits on the assumptions and machinery of a theory, they are not sufficient to provide one in and of themselves. In the next chapter, therefore, we develop our own theory through the addition of some further assumptions to the principles outlined above.

10 A theory of the septo-hippocampal system

In Chapter 1, we promised a theory of anxiety which would flow out of a theory of the septo-hippocampal system.[1] This might seem a retrograde step since the septo-hippocampal system, and hence any theory which purports to describe its operations, is complex. However, much of this complexity reflects the presence of parallel mechanisms, all of which serve the same class of purposes, and so the general thrust of our theory can be stated simply: the septo-hippocampal system resolves conflicts between concurrent goals, including, most importantly for our theory of anxiety, cases where conflict is generated by the presence of signals of impending punishment. The rest of the chapter will flesh out this summary statement, inevitably at the cost of some repetition of material that has gone before.

The septo-hippocampal system receives input from brain areas concerned with goal-directed action. If more than one of these is concurrently highly activated and no one of them is clearly predominant, the hippocampal system detects conflict. It resolves the conflict by returning to the active areas a signal which increases the relative valence of affectively negative information. Under conditions in which a weak approach tendency is due simply to weak positive motivation (for example, a lack of hunger), this signal will have no effect. However, if the weak approach tendency reflects the combination of strong positive motivation counterbalanced by previous negative associations (for example, with shock) of the same goal, the power of the latter associations will be increased. In consequence, otherwise prepotent behaviour will be inhibited and behaviour leading to the avoidance of negative outcomes will be favoured (Fig. 1.7). This process of conflict detection and increased behavioural inhibition continues, recursively and progressively, until the conflict is resolved and the animal approaches a goal. Where information in memory is insufficient to resolve the conflict (i.e. all current goals become suppressed), the hippocampus elicits information gathering (e.g. exploratory, risk assessment) behaviour.

On this view, the deficit experienced by an amnesic subject with damage to the hippocampal formation is the result of a failure to assign sufficient negative valence and hence to eliminate incorrect goals. On presentation of some retrieval cue, the subject will be faced (either explicitly in tests of recognition or implicitly in tests of recall) with a number of alternative goals. In the recall case, both specific retrieval cues and their context can activate a number of goal representations including those which would have been correct on previous trials. 'Behavioural inhibition' here is a filter which allows only one of the pandemonium of alternatives, each demanding attention, to control behaviour. In our theory, a balance between the strongest competing tendencies of any origin

1. To remind the reader, where the septo-hippocampal system is referred to as a whole, it includes the medial septum–diagonal band complex, the dentate gyrus, fields CA1–4 of the hippocampus proper, the subiculum, the entorhinal cortex, and posterior cingulate cortex; that is, the medial septum–diagonal band complex and all the areas which receive monosynaptic phasic inhibitory 'theta' input from it—see Appendices 5 and 6.

(including species-specific competing responses) will render any task sensitive to malfunction of the septo-hippocampal system, while a predominance of one goal over others will render a task insensitive, whatever the nominal psychological class of paradigm to which the experimenter supposes the task to belong.

The disturbance experienced in generalized anxiety disorder is the reverse of the consequences of hippocampal damage (see Chapter 11). The associations of stimuli which elicit a goal conflict are assigned excessive negative valence. In particular, cases of approach–avoidance conflict have the balance shifted in the avoidance direction. More colloquially (and in agreement with the data; see Chapter 11), the subject's perception of threat is increased. From this perspective, the principal clinical effect of anxiolytic drugs is to decrease the assignment of negative valence and hence decrease perceived threat. It is an important feature of our theory that this decrease will normally take some time to become fully effective, since previously acquired negative associations have, first, to be extinguished. The apparent immediate clinical effect of benzodiazepines probably represents a special case where temporary euphoriant and muscle relaxant actions mask the lack of a truly anxiolytic effect (Chapter 4, Section 4.1.6; Appendix 1). Note also that the effects of hyperfunction in the septo-hippocampal system will be evident only in cases in which there is in fact conflict, that is, a source of threat that has to be approached, not when active avoidance, motivated by pure fear, is engaged, as in simple phobias.

Despite more than 15 years of research since the first edition, the theory still reflects work very much in progress and is necessarily incomplete. Detailed mechanisms proposed in the first edition have been adjusted here to the new data which have become available over the last decade or more. We are sure that as many adjustments will be necessary in another 15 years. Important details are still unknown, such as the extent to which novel and classical anxiolytics differ because of the muscle relaxant and euphoriant side-effects of the classical anxiolytics or because of the corticosteroid-releasing side-effects of the novel anxiolytics; the precise number and type of gates within (and outputs from) the hippocampus; the behavioural functions of specific components of the system, such as the output from CA1; and, indeed, the detailed anatomy of crucial parts of the system, such as the ascending theta control network. As a result, the detailed mechanisms we postulate are often a matter of inference or guesswork, and a number of alternative specific schemes could probably be built of the same general type.

Nonetheless, in all key respects the theory is still fundamentally that of the first edition. The septo-hippocampal system is still, in Simonov's lapidary phrase, 'an organ of hesitation and doubt'. The primary business of the septo-hippocampal system is still to act as a comparator; although we now see this comparator function occurring at more levels than before. The most significant functional output from the septo-hippocampal system is still behavioural inhibition; although we have made the links with memory more explicit and increased the emphasis on information gathering. The septo-hippocampal system still represents the core of the behavioural inhibition system; although we have altered the precise weightings of some of the inputs (e.g. downgrading the medial septum and upgrading the dorsal noradrenergic bundle in relation to frustration). The septo-hippocampal system is, also, still the final common path through which anxiolytic drugs produce the majority of their effects.

The most significant alteration to the original theory is an addition. We now see the amygdala as an important direct site of action for the majority of anxiolytic drugs as well as an important route through which the 'increment arousal' and 'negative bias' outputs of the behavioural inhibition system are achieved. This allows a clearer statement of the nature of the partial overlap originally postulated between the septo-hippocampal system and the behavioural inhibition system: anxiety reflects the mutual interaction of the septo-hippocampal system and the amygdala. Activation of the septo-hippocampal system without the amygdala represents goal conflict resolution in the absence of the autonomic and arousal aspects (the gut feel, perhaps) of anxiety (and is likely to occur in pure memory tasks or approach–approach conflicts); activation of the amygdala without the septo-hippocampal system represents pure fear (among other possibilities), not the approach–avoidance conflict which underlies anxiety (see Chapters 2 and 3 for this contrast between pure fear and anxiety).

Overall, then, we offer a theory which has stood the test of time; which accounts for the role of the hippocampus in memorial tasks as well as do the most recently formulated memory theories; which (uniquely) accounts for the role of the hippocampus in non-memorial tasks; and which accounts also for the fact that one of the few known common neural actions of all the currently known clinically effective anxiolytic drugs is to impair septo-hippocampal function. Critically, the theory uses different patterns of activity within the same machinery to account for the involvement of the septo-hippocampal system in both memorial and non-memorial tasks and for certain specific pathologies of both memory and anxiety.

10.1 THE THEORY: SOME BASIC ASSUMPTIONS

In constructing the theory, we did not, of course, start with the ideas just outlined. We gathered many small pieces of information, considered many alternatives, and built up extensive dossiers which provided evidence for and against particular points of view. It has been like living in a detective novel, trying to decide who did it, or rather what it—the septo-hippocampal system—does. As in all the best detective stories, the real culprit will emerge only when we have constructed a theory to account for 'all' of the evidence. Let us begin where they always begin, by eliminating some of the suspects.

It is a good rule that the first suspect to come to the detective's notice is innocent. The first suggestion to be taken seriously was Milner's (1970, 1971) hypothesis that the hippocampus actively consolidates memories. However, as with a number of other memory theories discussed in detail in Chapter 8, this is not consistent with the nature of the animal and human deficits in recall or recognition after hippocampal damage. These now appear to arise from excessive storage of inappropriate information, rather than insufficient storage of appropriate information. From this we derive the first assumption for our theory.

Assumption 1 There is no long-term storage of information in the septo-hippocampal system. The apparent role of the hippocampus as an intermediate store (Rawlins 1985; Cohen and Eichenbaum 1993) is due to its interaction with other areas of the brain

where it normally ensures that incorrect items are not retrieved or stored concurrently with correct items during the intermediate phases of consolidation. This role of the septo-hippocampal system is not required once the requisite connections in these other areas have been consolidated, nor is it necessary for consolidation uncomplicated by interference. This view sits comfortably both with what is currently known about synaptic plasticity and with theories of memory that share Hebb's (1949) general approach to the neural basis of associative learning. Importantly, it views both a percept and the recall of the percept at all durations (including those appropriate for working memory) as resulting from activation of part or whole of a single-cell assembly. Thus different memories would be coded by the specific perceptual assemblies activated, and different levels of duration of memory would depend on whether the activity were supported by recurrent loops (as in working memory) or resulted from prior plastic changes in inputs to the assembly.

Another suspect that can be rapidly eliminated is Vanderwolf's (1969) voluntary movement hypothesis. Experiments on the behavioural correlates of theta activity in the rat hippocampus (Fig. 10.1; Appendix 6) suggest a strong relation between theta and movement through space. Since stimulation experiments (Appendix 7) show that theta is almost certainly an active state of hippocampal function, it follows that the hippocampus is doing something when the rat moves, consistent with Vanderwolf's proposal that there is direct hippocampal involvement in movement as such. However, the notion that theta, or its frequency, is necessarily related to movement must be abandoned, since not only low- but also high-frequency theta can be observed in unmoving animals (Sainsbury and Montoya 1984). The situation is also complicated by the fact that there appear to be two pharmacological gates, one apparently related to movement and one not, which determine (separately or together) whether the hippocampus is entrained by phasic impulses, the frequency of which is determined independently of the gates (Appendices 5 and 6; Fig. 10.2A,B). Worse, this frequency determination involves multiple, largely independent, sites in the hypothalamus (Fig. 10.2C). Furthermore, and critically, lesion experiments show that removal of the entire hippocampal formation has no obvious effects on the control of movement as such.

A similar discordance between correlation and lesion results occurs in single-unit experiments (Chapter 7 and Appendix 6). For example, if a stimulus achieves special significance for the animal (as when it is turned into a Pavlovian CS or the cue for an instrumental response), habituation of the hippocampal unit response is prevented; indeed, such stimuli appear to receive a privileged passage around the hippocampal circuit. But experiments reviewed in Appendix 8 demonstrate that precisely those tasks (simple classical or instrumental conditioning) which protect stimuli from habituation in the hippocampal circuit are immune to disruption by hippocampal lesions. This leads to our second assumption.

Assumption 2 The septo-hippocampal system receives and integrates much information about behaviour which it does not at that time control. More specifically, our theory postulates that the matching of subcortical with cortical goal information within the hippocampus indicates that an appropriate response is already programmed to occur (and hence there is no conflict) and so produces no output from the septo-hippocampal

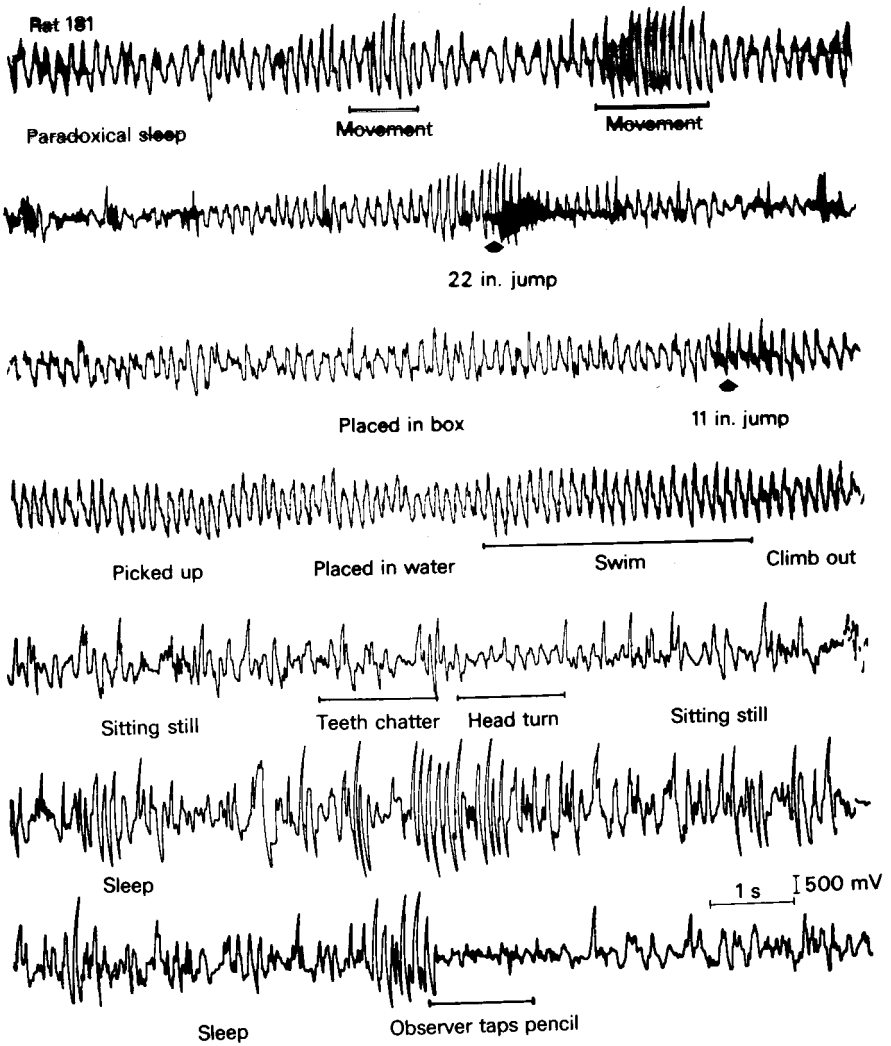


Fig. 10.1 Electrical activity in area CA1 of the hippocampus during sleep and various types of behaviour in the rat. Note the following: theta during paradoxical sleep, struggling when held in the hand, swimming and head movement; large-amplitude irregular activity during sitting while alert and while chattering the teeth; irregular slow activity and 'spindling' during slow-wave sleep and small-amplitude irregular activity when the rat was awakened but did not move about. Theta frequency increases with increasing movement (see the changes with twitching movements during paradoxical sleep and when making an 11-inch and, more so, a 22-inch jump) and theta amplitude increases with increasing vigour of movement (see the 11-inch and 22-inch jumps and swimming). Calibration: 1 s, 500 μV; half-amplitude filters, 0.3 and 75 Hz. (From Whishaw and Vanderwolf 1973.)

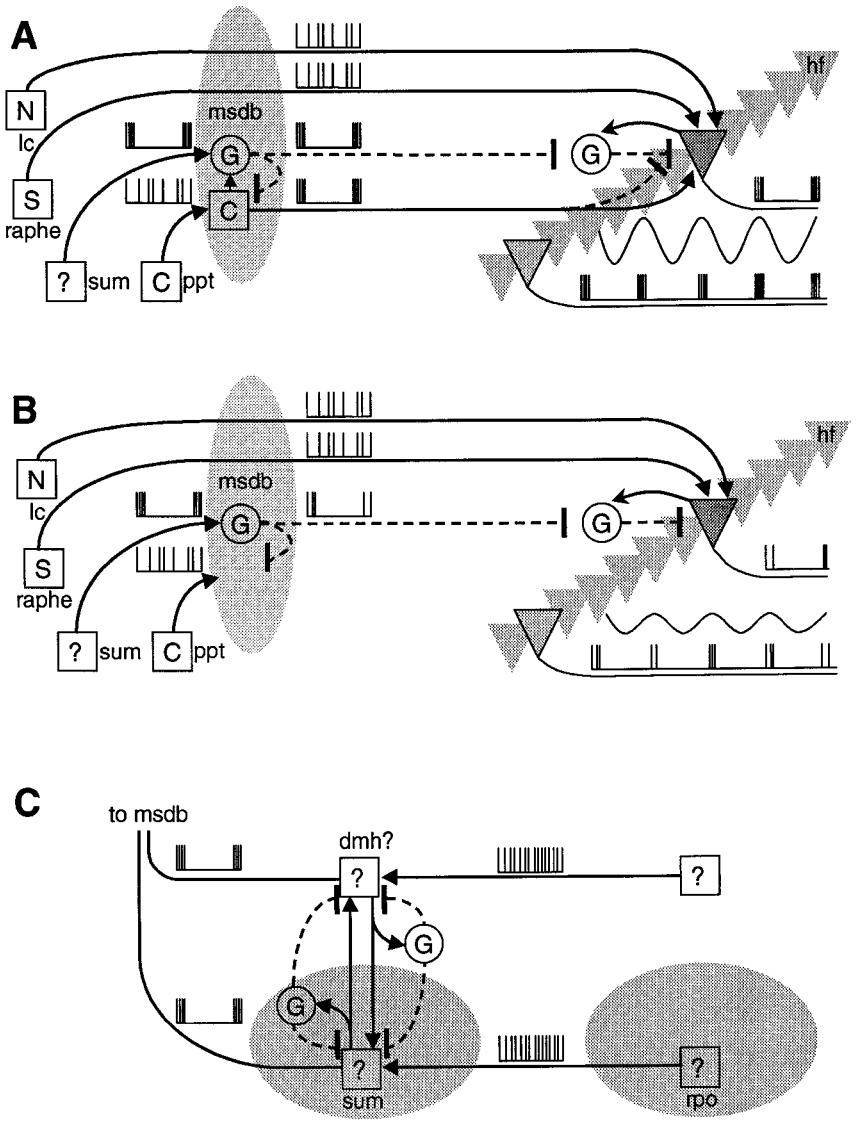


Fig. 10.2 Some of the circuitry controlling theta activity. (**A**) The medial septum–diagonal band complex (msdb) contains GABAergic (G) and cholinergic (C) cells which project to the hippocampal formation. The G cells are thought to synapse on G interneurons via inhibitory (- - -) connections. The C cells are thought to excite principal cells (filled triangles) of the hippocampal formation (hf) and to inhibit the G interneurons which, in turn, inhibit principal cells. Phasic drive of the msdb arrives via the medial forebrain bundle from areas such as the medial supramammillary nucleus (sum) and a cholinergic input which can gate theta activity arrives from an ascending cholinergic network of which the pedunculopontine tegmental nucleus (ppt) is part. (The transmitter for the sum–msdb pathway is not known, indicated by a question mark.) When the animal is stationary the phasic and cholinergic inputs are sufficient for phasic G and C input to be transferred to hf and so produce cellular theta activity (regular bursting action potential patterns shown

comparator. Much of the time, therefore, the hippocampal formation is busy 'just checking', but producing no functional output.

The elimination of our next suspect is more controversial. This is the theory (O'Keefe and Nadel 1978; Nadel 1991) that the hippocampus is primarily concerned with the construction and storage of spatial (or cognitive) maps. The spatial theory of hippocampal function offers a convenient way of accounting for the data on which Vanderwolf's voluntary movement hypothesis was based, since it allows the hippocampus to be doing something important (laying down and/or utilizing spatial maps) while the animal is moving, without requiring it to produce or inhibit movement. A vast amount of lesion data shows that, as predicted by the theory, spatially complex tasks are strongly disrupted by damage to the hippocampus. The spatial mapping theory is consistent with our Assumption 2, above, since it allows the hippocampus to be encoding spatial information even when there is no spatial requirement in the task, i.e. when *ex hypothesi* the hippocampus is not producing behaviourally effective output. However, the notion that spatial memories are actually stored in the hippocampus (recently defended by Nadel 1991) contradicts Assumption 1, above, although the theory could be reformulated to allow for permanent storage outside the hippocampus.

The spatial mapping theory also at first appears to receive strong support from the 'place fields' (whose discovery indeed inspired the theory's creation) seen in many of the single-unit studies of behavioural correlates of hippocampal activity. But a detailed analysis of those same single-unit studies (Chapter 7 and Appendix 6) suggests that it is very unlikely that spatial mapping (if it occurs at all) is more than one specialized aspect of more general functions of the septo-hippocampal system. Many cells have non-spatial correlates; and changes in the spatial fields themselves suggest that they are a consequence of the

---

schematically above the principal cell axons) and, in many areas, a gross extracellular hippocampal theta rhythm (wavy line shown to the right of the principal cell layer). Serotonergic (S) input from the raphe can also act to gate theta activity when the animal is moving so that, in this case, loss of the C input from ppt does not result in a loss of theta activity. Noradrenergic input from the locus coeruleus (lc) can alter the threshold for msdb activation of hf but does not appear to produce the more extensive gating effects of the S and C inputs. The interactions of N and S inputs with hf may be as complex as those proposed for C inputs but there are insufficient data to be sure. (**B**) Cholinotoxic lesions of the C cells of the msdb produce a reduction in the amplitude of theta rhythm. This effect is not produced by blockade of the ascending C system from ppt and is likely to be the result of a reduction of phasic activity in the G cells of msdb and the principal cells of hf. Amplitude can also be reduced by blockade of the output from sum. Amplitude of theta rhythm (and hence number of principal cells firing and/or their rhythmicity) can, therefore, be adjusted at each of sum, msdb, and hf. (**C**) The msdb receives phasic input from areas such as sum. High-frequency stimulation of nucleus reticularis pontis oralis (rpo) in urethane-anaesthetized rats sends tonic input to sum where it is recoded to phasic theta activity. Frequency is increased by any increase in afferent drive (by changing amplitude or frequency of stimulation) and is decreased by injections of benzodiazepines into sum. We hypothesize that frequency is encoded, therefore, by recurrent GAGAergic interneurons with GABA receptors linked to benzodiazepine receptors. There is evidence that in free-moving animals areas other than rpo can provide tonic drive which is transduced to theta frequency by areas other than sum and injections of benzodiazepines suggest that the dorsomedial hypothalamus (dmh) may contain one such area. There is also known to be some form of feedback from msdb or hf to sum, but that has been omitted from the figure. (Based on Lee *et al.* 1994; Vertes and Kocsis 1997.)

hippocampus receiving information about spatially correlated goals from other areas, rather than indicating in any unambiguous way the animal's location in space. Furthermore, lesions to the hippocampal formation disrupt behaviour in tasks from which all apparent spatial content has been removed. Good examples are Webster and Voneida's (1964) report of impaired extinction and reversal of a simultaneous tactile discrimination; the many reports of an extinction but not acquisition deficit in the Skinner box; and Gaffan's (1977b) experiment on the recognition of lists of colours. Damage to the hippocampus proper does appear to affect spatial tasks more than exteroceptive cue tasks (Jarrard 1991), but it also affects interoceptive cue tasks with no spatial component (Davidson and Jarrard 1993). So, in concordance with Squire and Zola-Morgan (1991), Cohen and Eichenbaum (1993), Gaffan (1994), and many others, we shall assume that spatial tasks contain some critical feature which renders them particularly sensitive to hippocampal lesions but which they share with at least some non-spatial tasks.

Assumption 3 The septo-hippocampal system plays no necessary role in learning, including spatial learning, as such. However (see Chapter 8), spatial discriminations, relational discriminations, and the like are particularly likely to generate conflict between explicit or implicit alternative goals (implicit alternative goals often constituting what we call 'interference', but also often involving species-specific reactions or other unplanned effects of particular experimental configurations); and it is the business of the septo-hippocampal system to eliminate this competition, where possible. Effects of hippocampal lesions will not be seen either where there is insufficient conflict or where the unlesioned control animals have been unable to overcome the conflict.

Our analysis rejects, then, memory consolidation, voluntary movement, and spatial mapping as complete accounts of septo-hippocampal functions. Alas, no other major suspect can be eliminated so readily. One problem is even to know how many suspects there are. Several appear to shade rather uneasily into one another: is sensitivity to interference different from loss of contextual labelling, non-reward from novelty, detection of familiarity from recognition memory, any of these from relational memory? Perhaps we are dealing with a master of disguises. Nor can we be certain that there is only one criminal—there could be a conspiracy. Non-reward or novelty might be responsible for the break-in, but behavioural inhibition for the break-out; and they might both have inside help (from attention, multidimensional stimulus analysis, or one of the other servants). Careful scrutiny of the suspects we have just listed suggests, indeed, that we could be dealing with a whole gang, of somewhat flexible membership, who usually work together rather than against each other.

When we consider this congeries of hypotheses, all more or less alike and all current in some form in the literature, a pattern recurs: they each explain some things well, but none can comfortably encompass more than a small portion of the data. Our fourth and central assumption, then, is one that lurks behind many of these theories, but which is not specific enough to constitute a theory in itself.

Assumption 4 The septo-hippocampal system acts as a detector and resolver of conflict between concurrent incompatible goals.

What precisely is meant by the term 'goal' is crucial to the theory and so, before we proceed, we should again make clear what this means. First, we see animals as learning about goals rather than learning to make responses. This is discussed in more detail in Chapters 1 and 3. But note, for example, that although well-learned responses may appear stereotypical, interference either with motor mechanisms or with the environment tends to result in the immediate appearance of new responses which achieve the same goal. Thus, changes which block the original response pattern do not initiate a period of new learning equivalent to the original learning 'of the response'. Second, a goal is the end-point of a response sequence and has no direct connection with any specific element of the sequence. Thus the response conflict which can occur in tasks such as mirror drawing (where there is only a single goal, but there is conflict between the expected and actual outcome of particular motor patterns) should not engage the septo-hippocampal system. Third, the concept of a goal has both stimulus and response aspects embedded in it (Fig. 1.7). Hence, goal conflict can result both from an apparent conflict between stimuli (as in a delayed matching-to-sample task, in which essentially the same motor response can be addressed to either of two physical stimuli) and from an apparent conflict between quite distinct classes of response which may be elicited by the same set of environmental stimuli (as in the classic approach–avoidance conflict). In both cases the animal must choose between two conflicting, mutually incompatible goals, and it makes little sense to suggest that we can ignore either the stimuli to which responses must be directed or the responses which must be directed to the stimuli. 'Goal' therefore combines both the stimuli and the requirement to respond.

10.2 THE THEORY: ARCHITECTURE

A diagram of the key components of the septo-hippocampal system and their proposed role in conflict resolution is shown in Fig. 10.3.

There are two basic conditions where conflict can occur. The first is when two incompatible goal representations are highly activated concurrently. Here, in essence, detection of conflict requires an integrator which sums inputs and produces output when the sum exceeds some threshold. The coding of response tendencies appears to be topographically mapped into the hippocampus (Appendices 4 and 6; Fig. 9.5), and so, at present, we view this integrator function as likely to depend on the fibres which carry information along the septo-temporal axis of the hippocampus. The second is when a prepotent response tendency is faced with unexpected circumstances: the occurrence of an unexpected event or the omission of an expected event. Here detection requires a comparator which produces an output when there is a discrepancy between observed and expected inputs. This can still be viewed as reflecting goal conflict, but not limited to a choice between two alternatives that remain the same from trial to trial. The unexpected event is likely to generate a number of potential response tendencies directed to a number of environmental stimuli. Both the response tendencies and the stimuli to which they are directed are likely to vary from trial to trial or from moment to moment. Furthermore, the precise treatment of each must be adjusted by the comparator in relation to the relative importance of the expected outcome of the specific action directed to specific

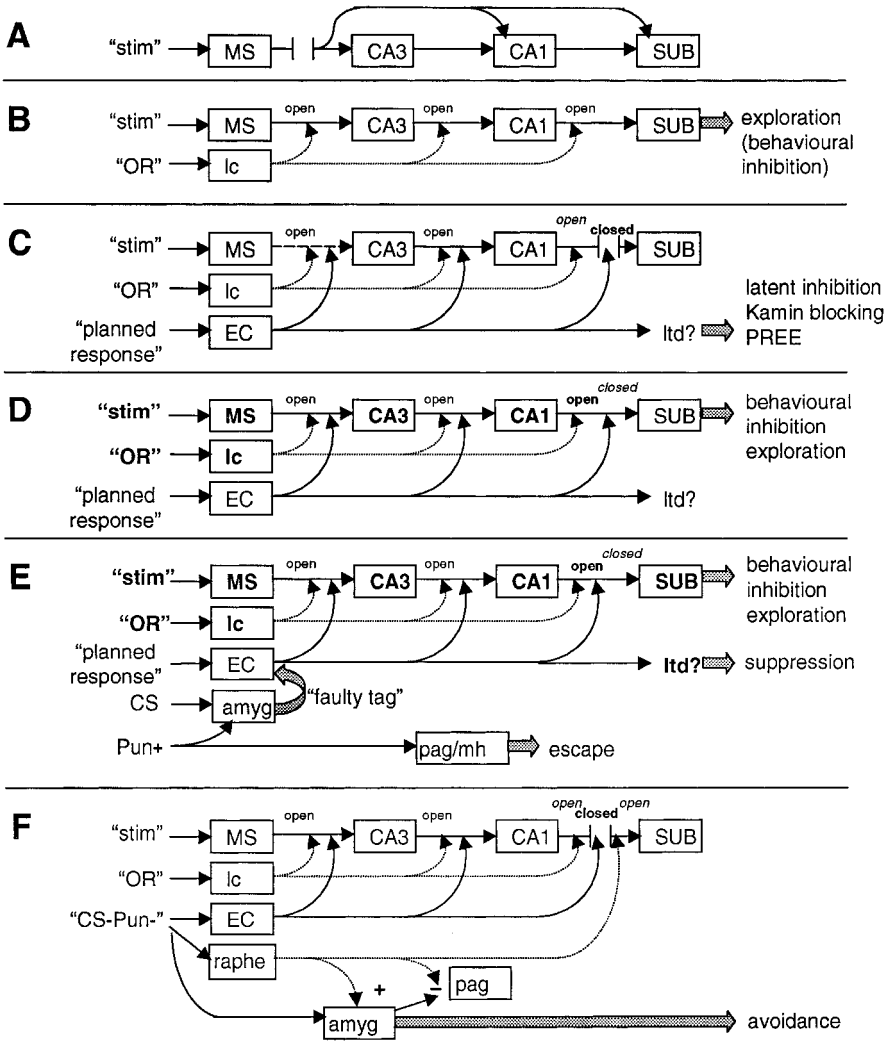
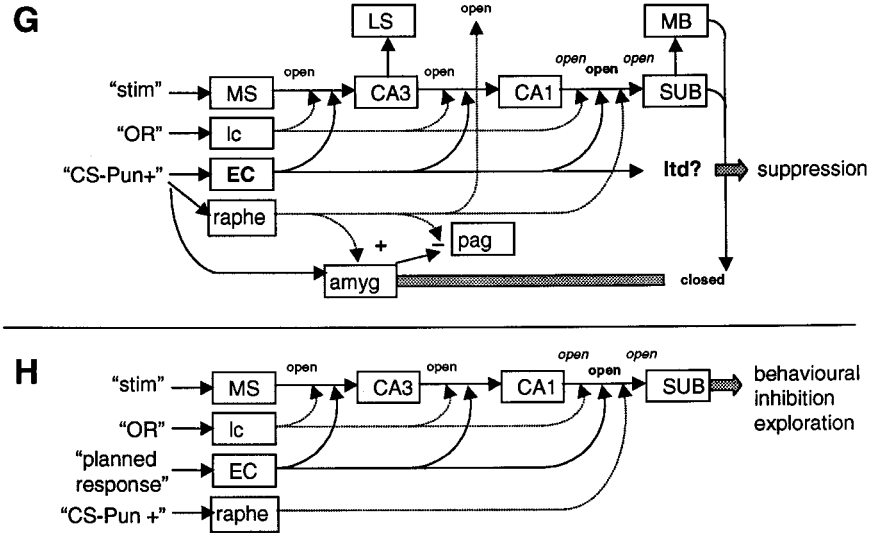


Fig. 10.3 The key components of the septo-hippocampal system and their role in conflict resolution (see also Fig. 9.4). For details see text, particularly Section 10.4. PREE, partial reinforcement extinction effect; MS, medial septum–diagonal band complex; CA1, CA3, subfields of hippocampus; SUB, subiculum; lc, locus coeruleus; 'OR', orienting–eliciting stimulus input; 'stim' sensory stimulus; ltd, long-term depression; EC, entorhinal cortex; amyg, amygdala; pag, periaqueductal grey; LS, lateral septum; MB, mammillary body. For definitions of learning theory terms (CS, Pun, etc), see Chapter 3. See opposite page for panels G and H.

environmental stimuli. As a result, in these cases a more stimulus-oriented description of the operation of the comparator will often be appropriate.

The hypothesis that the hippocampus acts as a comparator has been most extensively developed by Vinogradova (1975; see Appendix 6). According to her, the CA3 region of the hippocampus acts as a detector of novelty or of omission of expected events



through its reception of sensory input via the septum and of predictions of expected sensory input via the entorhinal cortex. But, as we discuss in Appendix 10, the lesion evidence makes it clear that the hippocampus does both less and more than detect novelty: less, because at the behavioural level the orienting reflex and its habituation are not disturbed to any great degree after hippocampal lesions; more, because the kinds of behavioural disruption which *are* seen after such lesions suggest that the hippocampal formation makes other comparisons besides simple judgements of familiarity.

However, if the hippocampal formation makes other and perhaps more complex comparisons than those required for the orienting reflex, it is unlikely that they are conducted exclusively in the CA3 comparator described by Vinogradova (1975). First, on her own evidence, the kind of information processing undertaken in area CA3 is insufficiently detailed to allow decisions that go much beyond 'novel' or 'familiar'. Second, all of the subfields of the hippocampus appear to receive equivalent information from both the medial septum and the entorhinal cortex. Thus, although the theory we shall develop here incorporates a version of Vinogradova's CA3 novelty comparator, it treats this as only part of a more extensive system which can produce output from additional comparators in CA1 and subiculum as well as CA3 (and which can also produce output from integrators in the entorhinal cortex and other areas such as the subiculum).

According to the theory, the function of the septo-hippocampal system is to detect conflict between goals, quite generally. It does so by indirectly monitoring information, not only about currently activated motor programs (necessarily directed towards specific goals and so directly represented in the hippocampus as such), but also about the stimulus representations that are at any time seeking to gain access to motor mechanisms. Such stimulus representations are also directly represented in the hippocampus as goals, but remain potential rather than actual at the motor level. In this monitoring capacity, the septo-hippocampal system functions continuously; but it controls other brain systems directly only under certain special conditions. Thus, it has two principal modes

of operation: 'just checking' (not controlling other brain systems) and 'control' (directly controlling other brain systems). These labels represent extremes of a continuum rather than distinct categories.

The principal function of the hippocampus proper in control mode is to narrow the focus of processing in extrahippocampal networks and to increase negative affective bias. An additional control function is provided by an instructed habituation/latent inhibition function of the entorhinal cortex (see Appendix 9). At some times, a small amount of control mode operation will be sufficient for adaptive behaviour; at other times, extreme amounts will be required. As with many emotional systems, we presume that continuous high levels of control mode activity are maladaptive, consistent with the fact that there are tasks where hippocampal lesioned animals produce superior performance to controls. (If this were not the case there would not be any need for the complex gating of information within the hippocampal formation. Further, we shall argue later, in Chapter 11, that the primary pathology in generalized anxiety disorder is excessive control mode activity of the hippocampus.)

The main part of the system we propose is what can be viewed as a second comparator, additional to Vinogradova's. This is described below; but it should be noted that the anatomy suggests that each output from the septo-hippocampal system may be from a separate (CA3, CA1, subiculum) comparator (or logical gate). The outputs from these comparators each address different subcortical, particularly hypothalamic, structures. They are also likely similarly to address different cortical structures. However, while we make specific assignments below, we cannot as yet unambiguously identify the function of any of these potential comparators, as there are virtually no relevant data. The specific details of this aspect of the theory remain, therefore, highly speculative.

The CA3 familiarity comparator is seen as auxiliary to the CA1 and subicular ones, sending on to them only signals that have special importance. The function of the subicular comparator is, so to speak, troubleshooting. It is called into play when the animal's normal routine is faced with other, conflicting, response tendencies. These latter can be elicited, notably, by novelty, the threat of non-reward, the threat of punishment, and relational processing. The task of the subicular comparator is to increase recursively negative affective bias in all of the active and conflicting goal-processing areas until only one alternative is clearly dominant. In this way existing plans can be applied again or new ones substituted as appropriate. Where there is insufficient information for resolution of the conflict to occur, all of the alternative goals will be suppressed. Under these conditions, output from the first, CA3, comparator will initiate general information-gathering behaviour (and particularly risk assessment); or output from the subiculum to the posterior cingulate cortex will initiate more complex forms of risk assessment, checking, and safety-seeking behaviour, including, for example, innate routines that do not depend for their termination on the achievement of a specific reinforcer. The conflict should then be resolved once the existing internal positive and negative affective associations have been updated with external information. These ideas directly reflect the psychological concept of the behavioural inhibition system developed in detail in Chapters 2 and 3. They also resemble closely everyday notions of anxiety and much of DSM-IIIR anxiety. But the purpose of the theory at this point is to account for the data on the septo-hippocampal system, not for anxiety.

The idea of conflicting goals is easiest to understand with respect to motor programs. The hippocampus is connected, directly or indirectly, to all major areas in which general goal-oriented motor programs can originate. In many situations the current stimuli (from both the internal and the external environment) will strongly activate only one such area and only one goal within that area. The hippocampus receives, as an efference copy, the information that a specific goal is being processed and this will activate a subpopulation of hippocampal cells. Since, in the case of a single such input, there is no conflict, the subject's motor behaviour can be released by the goal-processing area and will be directed towards the current goal. This is the 'just checking' mode of operation of the hippocampus.

Let us suppose now, however, that more than one goal is strongly activated. The hippocampus as a whole will receive much greater input than it would in the single goal case. (This could involve either a greater strength of total input to the hippocampus or a wider topographic spread of activation within the hippocampus.) Integrating these inputs, probably via intrinsic excitatory connections such as the longitudinal connections which run along its septo-temporal axis, the hippocampus will 'detect conflict' when the integrated sum effectively passes some threshold (in practice there may simply be an output which is proportional in strength to the degree of conflict detected). The output of the detector must then take active control of the subject's motor programs, either inhibiting one of the conflicting alternatives or inhibiting both and commanding appropriate exploratory or risk analytical action. These last might range from simple adjustments of autonomic or sensory organs (e.g. dilation of the pupils) to complex patterns of locomotion. In certain cases a compromise might be possible (e.g. the animal might continue to run towards the goal-box of an alley to obtain food, but do so more slowly while sniffing at the side-walls to gather information). Given the anatomical simplicity of the septo-hippocampal system and the complexity of the outputs we have just described, it is important to note, however, that 'inhibition' and/or 'exploration' are not postulated to be controlled in detail by the septo-hippocampal system as such. Instead, they are a consequence of the impact on other brain areas of the increases in negative bias which are produced directly by septo-hippocampal output.

An important feature of the theory is that, when conflict is detected, output from the hippocampus will influence not only current behaviour but also future behaviour by adding what is, in effect, a 'faulty' tag to one or more programs. Hippocampal output is most likely to achieve this tagging automatically, as an incidental though important consequence of the increase in negative bias through which it alters the balance between current goals. It is less parsimonious to assume that a distinct hippocampal output is required for this purpose. The inhibition of future behaviour is likely to depend upon inputs from the subiculum to the higher levels of goal-processing systems. In this way, the subiculum would be able to alter the affective bias of information arriving in specific goal-processing systems from, for example, polymodal association cortex. Again, the complexity of the processing just described is presumed to reside not in the hippocampus itself but in the responses of its many output targets (some active, some not) to a modulatory signal from the septo-hippocampal system, a signal that might itself have no target specificity at all.

Note also that the final effect of inhibition of a specific prepotent appetitive goal is likely to be the result of hippocampal excitation of a competing aversive goal. The greater the competition and the greater the hippocampal output, the greater will be the arousal

and attentional concomitants of hippocampal activity. Critical for the anxiety-related aspects of the theory is the idea that hippocampal excitation of specific nodes will result in long-term potentiation in extra-hippocampal targets that would not otherwise have occurred; and that this in turn results in subsequent retrieval of affectively negative (in the approach–avoidance case, 'threatening') associations at greater strength than would have occurred in the absence of functional hippocampal output.

The 'faulty' tagging in memory must involve an asymmetry between reward, on the one hand, and punishment, on the other. If the current motor program is interrupted by an unexpected reward, it may need to be modified; but the plan elements could be retained until the reward was obtained. However, if it is interrupted by an unexpected punishment, it is likely that the whole program should be treated as faulty. Thus, the conflict detector needs not only information about the presence of conflict, in a purely cognitive sense, but also information about the general affective value of unexpected stimuli.

The capacity to resolve conflict between two incompatible goals which we ascribe to the hippocampus is, then, inherently very simple. The hippocampus is functionally activated simply by increases in the topographic spread of the inputs to it from cortical and subcortical-goal processing areas, and so requires little in the way of specific complex information. This is consistent with the fact that the hippocampal formation is phylogenetically very ancient. In animals such as fish it represents a major part of the cortex (Gastaut and Lammers 1961). It is unlikely that sharks, for example, have as great a need for relational (or whatever) memory as do human beings, but easy to believe that they have as great a need to resolve conflicts between concurrent incompatible goals (since these can as well be innate as learned).

This model of detection and resolution of conflict between goals, in which the goals have so far been assumed to differ in the nature of the response to be made (e.g. approach, avoidance), can be applied equally well to the resolution of conflict between goals that differ in the stimulus to which to direct a particular type of response (e.g. a choice between levers to press). There will be conditions where two stimuli, assigned as positive and negative by the experimenter, might appear to be equally attractive goals to the subject (depending on the nature and history of the stimuli and of the contexts in which they have been presented). On detecting conflicting items, each priming a response system, the hippocampus will selectively dampen activity (via increased negative bias in both structures) so as to leave only one item available for transfer to working memory. It will also tag incorrect retrievals, so that their probability of occurrence in a similar context is reduced in future. There appears, however, to be some differentiation in the hippocampal formation, with the entorhinal cortex being more stimulus-oriented and the hippocampus proper being more response-oriented. We would expect such differentiation to become clearer and richer as the details of the topographic organization of the hippocampus (e.g. Fig. 9.5) are worked out. It is also possible that theta rhythm and hence anxiolytic action are more significant for hypothalamically mediated 'response–response' conflicts and less significant for cortically mediated 'stimulus–stimulus' conflicts.

We should also note the special case of novel neutral stimuli. Here the stimulus itself need not be an innate elicitor of some incompatible response. Nor does the stimulus necessarily conflict with a second stimulus for access to the upcoming response. Rather, the presence of the stimulus (independent of the response it elicits), since it is unexpected,

has the capacity to inhibit any ongoing response. To the cases of nominal response–response and nominal stimulus–stimulus conflicts considered in the previous paragraph we must, therefore, add nominal stimulus–response conflict. All of these, however, can be recombined as different flavours of goal–goal conflict. In particular, according to the theory, a novel stimulus will only produce functional output from the hippocampus if it is not merely novel but is also the goal of an orienting response.

While we have just described a number of different types of conflict with which the hippocampus is supposed to deal, these cannot be subsumed under the heading of 'conflict' in general. In the theory presented here the septo-hippocampal system is responsible only for the detection of conflict between concurrently activated goals. Conflict between successively activated goals (essentially ordering) is presumed to be dealt with in the prefrontal cortex (Chapter 6; Appendix 3); and conflict between different motor patterns which can achieve a single goal is presumed to be dealt with by the motor system (see also discussion of the defence system in Chapter 6), particularly the basal ganglia.[2] This is not to say that these different conflict resolution systems are totally independent. Cortical systems, including in particular the prefrontal cortex and anterior cingulate cortex, furnish predictions about upcoming events or actions, many of which are funnelled through the entorhinal cortex. They are, therefore, essential to at least a subset of hippocampal operations (as it is, in turn, essential to a subset of their operations).

10.3 THE THEORY: MECHANISM

We may seem in the above to have ascribed a fairly large number of quite complex functions to what is in many ways an anatomically simple structure. However, the specific computations we assign to the hippocampus are simple; the complexity of the resultant functional output at the psychological level is a consequence of the response to hippocampal input of its target structures. In this section we consider some of the specific operations carried out by the hippocampal formation.

The most fundamental operation, in our view, is recursion: the 'recycling' of activity back to structures which are, in essence, the source of that activity. This 'recycling', however, is not of 'the same' information. At each cycle the information is modified not only by the consequences of hippocampal input for the relative weighting of affectively different classes of information in the source of the original information, but also by the current state of the hippocampal circuitry.

The functioning of this basic recurrence in hippocampal connections can be understood by analogy with the global stereopsis and language models which we considered in Chapter 1. The hippocampal formation as a whole appears to do a great deal of 'talking to itself', that is, it sends out and gets back messages via the systems with which it is connected reciprocally (systems whose higher levels receive, for example, subicular output, and whose higher and lower levels return information to, for example, the entorhinal cortex); and, since it sends its message out by various different routes, it also gets it back again (modified) in multiple copies. We must also account for the fact that

2. We are grateful to Dr Lynette Tippett for pointing out to us the necessity to make these distinctions clear.

within the hippocampus proper, by contrast to its extrinsic connections, there is an essentially linear, unidirectional flow of information; and must bear in mind that, as far as the septal and entorhinal inputs are concerned, all levels of the hippocampal linear array appear to receive essentially the same information in parallel. In our view the purpose of this architecture is to permit the proper functioning of a series of logical gates. Each receives much the same septal and entorhinal input as the others, but they perform different calculations on this input, integrating in particular the septal and entorhinal inputs with other signals, especially those from the ascending aminergic systems.

The major terminal point of this cascade of processes within the hippocampus must be the subiculum, as this is not only the final point up to which unidirectionality is largely maintained (output from the posterior cingulate is generally reciprocated by the areas to which it projects), but is also a major source of efferents from the hippocampus. Information reaches the subiculum from CA1 after passing around the basic hippocampal circuit, i.e. from the entorhinal area to the dentate gyrus via the perforant path, thence to CA3 via the mossy fibres, and on to CA1 via the Schaffer collaterals. It is then recirculated through, for example, the fornix to the mammillary bodies, thence to the anteroventral thalamus, the cingulate cortex, and back to the subicular area. (There is also a smaller loop providing direct two-way connections between the subiculum and anteroventral thalamus.) Given this architecture, the subiculum is a prime candidate for the role of goal conflict detector (in contrast to the mismatch comparator in CA3). It receives general information ('arousal') via the septum and aminergic inputs; information that predicted stimuli (goals) have occurred via the entorhinal cortex; details of motor programs/working memory (goals) from the prefrontal cortex; and details of innate programs (goals) from the cingulate cortex, amygdala, cerebellum, etc.

The subiculum is also positioned so as to be able to interact recursively with all these areas (e.g. subiculum–mammillary body–anteroventral thalamus–cingulate cortex–subiculum) so as to modify their function. Such modification (if it is based on phenomena like long-term potentiation and depression) must proceed progressively round the loop from the subiculum (see below). There is some evidence that this is the case. As we note when discussing the cingulate cortex in detail (Appendix 3), single units recorded during discriminated active avoidance react differentially to S+ and S– during acquisition in both the anteroventral thalamus and the cingulate. Behavioural acquisition of reversal of the task is accompanied by incomplete reversal of unit responses in the thalamic neurons (which are closest to the subiculum) and by little change in the discriminative responses in the cingulate cortex. Thus, the hippocampal system does not eliminate previously formed associations, but alters their capacity to control behaviour.

If simple detection of conflict between alternative courses of action, independent of type, were the basis for hippocampal function, then perhaps direct input of information about those courses of action to the subiculum would be all that is required. However, both behavioural data and the general theory developed in this book require us to treat presentation of a punisher differently from presentation of a reward. The theory needs, in particular, to suppose that the omission of an expected punisher can produce, in the hippocampal circuitry and particularly in the subiculum, an effect equivalent to that of the occurrence of an expected reward, i.e. leaving the system in the 'just checking' mode. We propose that differing treatments are accorded to conflicts of different types as a

result of the operation of a series of gates between the different hippocampal subfields. The anatomy suggests that essentially the same neural messages, representing the animal's current plans, can be sent all round the hippocampal circuit to the subicular area, as well as directly by way of the entorhinal-subicular projection (Van Hoesen *et al.* 1979). We propose that these inputs interact with what are in essence logical gates (Fig. 10.3) so that, in some cases, the subiculum will receive only a direct input from the entorhinal cortex, while in others it will receive parallel input both from the entorhinal cortex and, as a result of specific gates being open, from the hippocampus proper.

It is in the four synapses of the, superficially unnecessary, unidirectional loop through the hippocampus proper (Fig. 10.3) that Vinogradova (1975) places her proposed basic comparator function. The single-unit evidence, both her own and that of others, as well as the research which has followed evoked potentials round the hippocampal circuit (Appendix 5), gives strong support to her hypothesis that this hippocampal circuitry can show habituation to, or preserve from habituation (depending on the importance of the message), signals emanating from the entorhinal cortex and destined for the subicular area. The main purpose of this circuitry (we propose) is to filter information destined for the subiculum, so that the latter receives only those items of information which need to be assessed for conflict value. As detailed in Appendix 5, there appears to be a logical gate located between the dentate gyrus and CA3. This passes on stimuli which are novel (i.e. for which there is no coincidence between septal and entorhinal input), but not those which are both familiar and lacking in behavioural significance (Vinogradova 1975). However, as a result of the operations of the ascending monoamine systems (Appendix 10), familiar stimuli are also passed on if they have been associated (by classical conditioning) with events of primary biological importance, or if they are of conflicting significance. Segal's (1977a,b) experiments illustrate this point well: with the parameters of electrical stimulation on the entorhinal side of the dentate–CA3 gate set so that responses on the other side rapidly habituated, either concurrent presentation of environmental CSs associated with punishment or direct stimulation of the locus coeruleus prevented this habituation.

Note, however, that the phrase 'passes on information' in the above formulation is imprecise. Vinogradova's experiments make it very clear that the information which is processed in CA3 is of a very non-specific kind. As she puts it 'the neuronal reactions in CA3 do not code (and consequently cannot transmit) information about the quality of their sensory input. Their activity appears to be a strong generalized regulatory signal, which may possibly exert modulatory effects on the output structures' (Vinogradova 1975, p. 10). The dentate–CA3 gate must work, therefore, in parallel with the direct pathway from the entorhinal area to the subiculum. Thus, specific information is sent continuously to the subiculum, but it only affects processing in the loops connecting the subiculum with other areas if it receives an 'enabling signal' from CA1. This enabling signal is sent only in connection with combinations of inputs that signify conflict, otherwise they are filtered out during their passage through the dentate–CA3 gate. Thus, direct input from the entorhinal area to the subiculum describes the state of the world, while the input relayed via the other subfields of the hippocampus, utilizing the gating capacities illustrated in Fig. 10.3, determines the precise kind of comparison which is carried out, depending on the way in which the stimuli are important to the animal. If

one views the hippocampal circuits in this way, then the known effect upon them of monoamine inputs (Chapter 9) could be phrased in English as a tag stating 'the current entorhinal input is important, it needs checking'.

The entorhinal area is probably the major funnel for cortical information (Fig. 10.4) arriving in the subiculum, and appears to be the only funnel for cortical information to

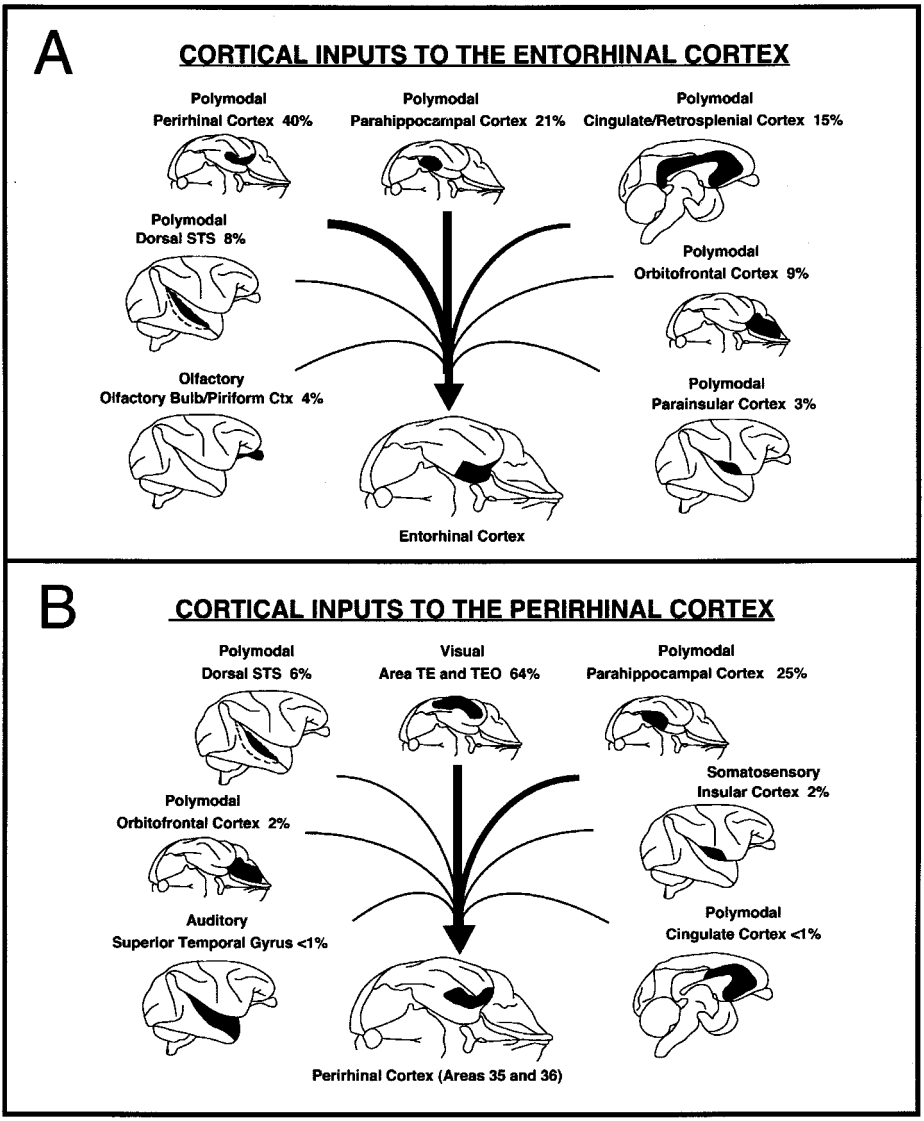
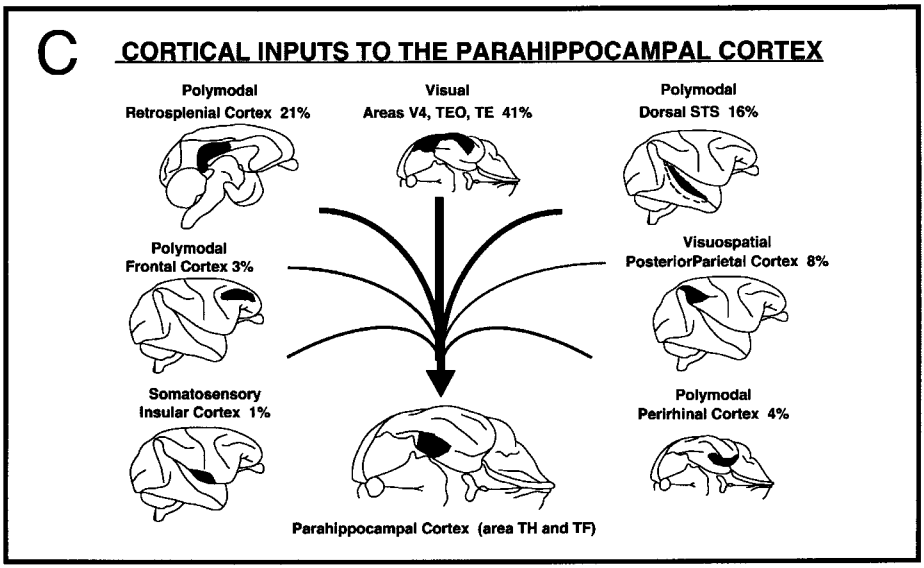


Fig. 10.4 Cortical inputs to the entorhinal cortex (A) are largely relayed by the perirhinal (B) and parahippocampal (C) cortices. Density of connections is indicated by width of arrows. STS, superior temporal sulcus; TE, TEO, V4, eponymous areas of visual cortex. See opposite page for C. (Figure courtesy of W. A. Suzuki, data summarized from Suzuki and Amaral 1994; Insausti *et al.* 1987.)



other areas of the hippocampus in non-primates. However, in primates there are inputs to subiculum, as well as to other parts of the hippocampus proper, directly from prefrontal cortex, parietal cortex, temporal cortex, and cingulate cortex. Such direct inputs could fulfil the same general functions as the information from those areas which is normally relayed by the entorhinal cortex. While our above argument, therefore, has been phrased in terms specifically of entorhinal cortex (and in the absence of many further data should continue to be so), it can be applied with no functional distortion to any route by which cortical information reaches the septo-hippocampal system.

In this context, we should recall the type of information which arrives via the septum. We have already alluded to the fact that cellular responses in area CA3 are too broad to code for the external stimulus features of the world which the subiculum will require for conflict resolution. This problem is not so severe if we assume that the information is coded in general 'goal' terms; but, even so, the pathways reaching the hippocampus via the septum are far too modest to code highly specific information of the type needed for conflict resolution. In the same way that we viewed CA3 as providing a general signal which then determines whether or not the subiculum processes information, it seems reasonable to suppose that the information arriving from the septum codes for quite general properties (perhaps simply 'importance', as indexed by intensity of activation), and that other routes provide specific information as to the nature of the goal giving rise to this alerting signal. This is in keeping with both the reticular origin of the septal information and the indiscriminate integration of activity apparently carried out by the theta-frequency-coding supramammillary nucleus and the theta-gating ascending cholinergic system (Fig. 10.2; Appendix 5 and 10). Thus the input from the septum can be seen as classifying information as important in the same general way as do the monoamine inputs (a view consistent with the similar effects of acetylcholine,

noradrenaline, and 5-hydroxytryptamine on neuronal signal-to-noise ratios), but on a dimension of potential response intensity rather than the dimensions of response quality (determined by the nature of reinforcement) signalled by the monoamine inputs.

It is this input from the septum which makes habituation of hippocampal responses an active rather than a passive process (Vinogradova 1975). As described in Appendix 6, disruption of either the entorhinal or the septal input to CA3 prevents habituation of the neuronal responses in this region. From this evidence it can be deduced that hippocampal habituation depends on some kind of matching of these two inputs in CA3. If we consider Vinogradova's findings in more detail, it seems that the changes which give rise to hippocampal habituation take place on both the septal and entorhinal sides of CA3, but in rather different ways.

On the entorhinal side, there is a gradual build-up of responses to repeated, unimportant stimuli in both the dentate gyrus and the entorhinal cortex itself. The build-up of response in these regions takes about the same number of stimulus presentations (2–12 in the entorhinal area, 15–20 in the dentate gyrus) as does habituation in CA3 (8–15). This is consistent with the fact that long-term potentiation of the entorhinal input to the dentate gyrus blocks sensory responses within the hippocampus. Thus, long-term potentiation of this kind has an effect (loss of response in CA3) that is equivalent to instant habituation of responses to most neutral stimuli simultaneously. In parallel with the spread of incremental responses from entorhinal cortex to dentate, culminating in habituation in CA3, there is a spread of incremental responses progressively from CA1 to mammillary bodies to anteroventral thalamus to cingulate cortex (Fig. 10.6), with maximal augmentation being obtained in the cingulate in 8–15 trials, concurrent with habituation in CA3. In line with our arguments earlier, the cingulate could then be viewed as holding at least a partial model of the occurrence of the stimulus.

This formulation places the burden of hippocampal habituation on the entorhinal side of the CA3 gate. However, the septal area also plays an active role in the process. Unit responses in the medial septal area neither augment nor show much habituation. Lateral septal unit responses undergo habituation, which appears shortly after habituation of CA3 neurons, and so is more likely to be a consequence than a cause of CA3 habituation. Sectioning the fornix-fimbria disrupts normal habituation of CA3 responses and replaces it with augmentation (Vinogradova 1975; Vinogradova and Brazhnik 1978). This pattern of effects suggests that what we are seeing is not what is conventionally thought of as habituation at all (remember that hippocampal lesions do not interfere with the habituation of behavioural-orienting responses to the stimuli used in these experiments). Rather, like the aminergic inputs from the subcortex, the septal input increases signal-to-noise ratios, doing so on the basis of average stimulus intensity across a range of modalities. This is consistent with the fact that theta frequency increases with increasing midbrain activation essentially independently of the way in which that activation is produced, with the supramammillary nucleus functioning as a simple integrator.

The phasic input from the septum to CA3 differs from the aminergic inputs, however, in the precise way in which it alters signal-to-noise ratios. The septal input is predominantly GABAergic and inhibitory, depressing presynaptic release of neurotransmitter in many terminals at the same time that the excitability of the cells is increased by other (e.g. cholinergic) inputs. Thus the septal GABAergic input changes signal-to-noise ratio

largely by suppressing noise. In the absence of the septal input, depolarization may be sufficient to induce long-term potentiation, particularly under conditions which produce hippocampal sharp waves; these, according to Buzsaki and Haas (1988; see Appendix 5), promote long-term potentiation (theta having the opposing effect). However, in the presence of the septal input, the majority of other concurrent inputs will be taken below the critical depolarization threshold and undergo long-term depression. In contrast, in cases in which there is significant input from the monoamine systems, signal-to-noise ratio is increased largely by an increase in signal, and the capacity for long-term potentiation, at least of the most active concurrent inputs, is restored.

Thus, the hippocampus consists of a cascade of logical gates, each of which can, but need not, produce functional output. The interaction of these gates, and their modulation by ascending aminergic inputs, then allows not only for the distinct modes of operation of the system ('just checking', 'control') but also for quite different logical combinations of simple stimuli and quite different resulting states of the goal-processing systems of the brain, all nonetheless to be treated as sharing in the higher-order property of 'conflict'.

10.4 SOME SCENARIOS

In order to see how this theory (including the major assumptions detailed above) can be applied to the available data, it is necessary first to distinguish between four largely different general scenarios.

Scenario 1 The animal is in a totally, or significantly, new environment. Under these conditions activation of the subcortex will result in septal input to the CA3 comparator which will not be matched by a predictive cortical model of the environment. The mismatch between these inputs will signal potential conflict between different possible courses of action. As a result the septo-hippocampal system will tend to inhibit all prepotent goals and elicit the gathering of information which will allow appropriate choice of subsequent goals. Where stimuli are not linked to important changes in reinforcement contingencies, the formation of a cortical model of those stimuli (which could take the form of a 'null' motor program) will result in a 'familiar–ignore' signal being generated.

Scenario 2 There exists a set of continuously updated expectations which are continuously matched by current sensory input. Under these conditions, according to the theory, the CA3 comparator receives matching subcortical and cortical signals and will exhibit cellular firing which reflects this fact, but, since they match, exercise no control over behaviour.

Scenario 3 The continuously updated expectations fail to be matched by actual events. This includes the case where the only deviation from prediction is that some event fails to occur. Both unexpected presence and unexpected omission can be viewed as variants of Scenario 1. Under these conditions the septo-hippocampal system assumes direct behavioural control. Since this will usually occur in the context of an ongoing motor program, a major feature of the control is the tagging of the prepotent, currently executing,

motor program as 'faulty needs checking' (leading to subsequent as well as current behavioural inhibition). Concurrently, there is activation of information-gathering behaviour aimed at resolving the discrepancy that has emerged between expectation and outcome. As noted above, appetitive and aversive stimuli will be accorded somewhat different treatment, with aversive interruption requiring more extensive motor program modification. As in Scenario 1, whenever the interrupting stimulus (including omission of reward and other affectively loaded stimuli) is unrelated to subsequent reinforcement, the stimulus will eventually be tagged as 'familiar–ignore', and the initial 'faulty' tag assigned to the active motor program will be removed. (This gives rise, among other things, to latent inhibition and the partial reinforcement extinction effect; see Appendix 9.)

Scenario 4 After the discrepancy has been resolved, behavioural control passes back to other systems which may, however, receive updated information as a result of the activities of the septo-hippocampal system.

We now consider each of these scenarios in more detail.

10.4.1 **Scenario 1: exposure to a novel environment**

We treat this scenario (see Fig. 10.3B,C,D) first because it represents the first step in any dealings which the animal will have with any particular stimulus. With an adult animal, however, Scenario 2 can be viewed as being the 'normal' or 'baseline' situation.

We have already commented on the lack of change in simple orienting responses or their habituation after septal or hippocampal lesions. Leaving aside the frequently observed phenomenon of motor disinhibition in novel environments, the few reports of a slight retardation in habituation or of a reduction in the galvanic skin response in lesioned animals (Appendix 8) can hardly be considered evidence of a fundamental disruption in the intake of information about novel stimuli, as required by Vinogradova's original hypothesis. There is in any case no need to postulate an entire organ of the brain which supports habituation, since in simple organisms, such as *aplysia californica,* habituation, sensitization, and association can all occur within monosynaptic reflex arcs (see for example Pinel 1997, p. 390). Yet, at the same time, it is clear from the EEG and single-unit studies (Appendix 6) that the hippocampus is engaged at such times in registering novel events and, if they have no special significance, in registering this lack of significance as well, by habituating the responses they evoke.

This largely 'hands-off' reaction to novel, and then less novel, information can be better understood if we recall the details of eye-blink conditioning. The single-unit data make it clear that the associational connections made during such conditioning are reflected in changes in hippocampal unit activity. However, the lesion data make it clear that the septo-hippocampal system plays no part in the acquisition of the conditioned nictitating response. Rather, this is thought to occur in the cerebellum. Thus, the hippocampus receives efference copies from circuits which control simple associative learning, but the hippocampal cells themselves cannot (and need not) distinguish easily between the areas supplying these efference copies, and only take control of behaviour under conditions which these basic circuits cannot handle. As noted in Chapter 6, a

similar hierarchical organization appears to be present in the threat system. It is parsimonious, therefore, to presume that the same principles operate in relation to habituation to simple novel stimuli, a case in which it can again be supposed that the circuits required to process the relevant information are extra-hippocampal. As with changes in hippocampal responses during simple eye-blink conditioning, then, we may suppose that the hippocampus is not totally unconcerned about the information it receives about novel stimuli. While it is not functionally involved in the habituation of the orienting response as such, nor the storage of the cortical or subcortical model which underpins habituation, the changes in its unit responses are important for the role it may play in more complicated future tasks (as in the parallel cases of trace and delayed eye-blink conditioning).

The assumption that the hippocampus does not store specific details of the information passing through it (and hence cannot code this information for use by its output structures) is supported by the rapidity with which hippocampal units picked out at random by the experimenter's electrode respond to stimuli in a totally new environment (Chapter 7 and Appendix 6). This could not happen if the units were already dedicated to the storage of specific information. Observations of this kind are, in contrast, consistent with the idea that the hippocampus continuously integrates information arriving from other cortical (and subcortical) areas.

This is not to say that there are no disturbances of behaviour in novel situations in animals with septal or hippocampal lesions. There is a noticeable disruption of the patterning and flexibility of active exploration of spatially extended environments. (Unfortunately, there appear to be no studies of active exploration of complex, but non-spatial, environments.) A particularly marked change is the loss of rearing. This is a major component of the risk-analysis behaviour seen in response to anticipation of predatory threat (Chapter 2). In a novel environment, rearing could reflect similar activity in the threat system; but it is just as likely that the multitude of available stimuli elicit conflict which, even in the absence of specific threat, requires the same general investigation of the environment.

Some novel environments involve an implicit approach–avoidance conflict, of the kind that is central to our analysis of anxiety. The approach component is based on the advantages attached to exploration (such as discovery of new food sources); the avoidance component is based on the potential threat from as yet undiscovered environmental dangers (such as predators). In our view, the hippocampus is not involved in dealing with novelty or its habituation *per se*, but it does take control of behaviour when a novel environment elicits conflict between a high degree of activation of approach and a simultaneously high degree of activation of avoidance. Even in a spatial environment, by the way, it is not the case, as O'Keefe and Nadel (1978, p. 242) suggest, that 'in the absence of the hippocampus all forms of exploratory behaviour . . . disappear from the animal's repertoire'. The lesioned animal still explores and (in a straight alley for example) learns its way about and identifies the location of significant stimuli. The chief difference from normal behaviour is that the pattern of exploration is less flexible, less well adapted to ensuring that all parts of the environment are visited. We shall consider later how these effects can be explained within the present theory (see Scenario 3).

In sum, during processing of novelty, the hippocampus is not directly involved with the construction of predictive models of the stimulus situation, nor directly with

production and habituation of the orienting reaction. Rather, it receives subcortical input which reflects the presence of important stimuli (potential goals). If the hippocampus does not at the same time receive a matching cortical input (i.e. the goals are novel), it determines the relative strength of the novel goal and of any prepotent goal; and, if there is significant conflict between these (i.e. neither is significantly greater than the other), then it produces an output which inhibits the prepotent goal and hence permits the orienting and exploratory programs to function properly. This is not a dichotomy, in that intermediate cases (with more or less slowed execution of a prepotent response intermixed with more or less exploration) are possible.

10.4.2 **Scenario 2: just checking**

In this scenario (see Fig. 10.3A,C,F), the cortex and subcortex provide a continuous stream of information about present goals and predicted goals (plans). There are two varieties of this scenario. In the first, the animal does nothing much to change its environment and change, or lack of change, in the environment simply lives up to expectation (as, for example, in a habituation or classical conditioning experiment). This is the natural end-point of the cortical model-building elicited during Scenario 1. The second variety is more interesting and has been subjected to more empirical analysis, so we shall take it for our principal object of discussion. Consider, as a concrete example, a rat running in an alley for a food reward on a simple continuous reinforcement (CRF) schedule. That the hippocampus is active under these conditions is indicated by the continuous high-frequency theta which is observed as the rat runs to the food cup; and the data from single-unit experiments show that many cells will be firing, each in relation to its own spatial location field (O'Keefe and Nadel 1978); according to our theory, this pattern of firing will be determined by the available goals (see Chapter 7).[3] Yet we know that a rat without a septal area (and hence without a theta rhythm), or even a rat without a hippocampus, runs just as efficiently as an intact animal. Indeed, particularly in operant chambers, it may even respond more efficiently than a control. In this situation, then, the hippocampus is 'just checking'. Its units are receiving and integrating multiple sources of information, but the operation of the logical gates in the hippocampus is such that little functional output occurs. Where output does occur, this reflects minor discrepancies between observation and expectation (or between action and consequence) which result in inhibition and exploration. It is the loss of processing of such discrepancies which sometimes produces an apparent improvement in learning after hippocampal lesions.

It is worth considering in more detail how this checking occurs. Following the arguments outlined above, the hippocampus receives information from all areas which could initiate motor programs, and all areas which analyse the complex stimulus configurations on which such programs depend. This information is received by the hippocampal formation (particularly the entorhinal cortex) from (for example) the cerebellum, the

3. One might prefer the term 'subgoals', if one wished to restrict the meaning of the term 'goal' to the food cup. But this should be avoided, as a major problem in this literature has been to assume that the experimenter knows and defines the rat's goals, via the paradigm, rather than the rat itself determining what is or is not a goal—something that must then be discovered by the experimenter.

amygdala, the anterior and posterior cingulate cortices, the dorsal and ventral trends of prefrontal cortex, and from a wide variety of sensory areas, particularly polymodal association cortex (Appendix 4). All of this information must be integrated to determine the degree of conflict, i.e. the degree of similar coactivation in different areas. In the simplest case, the hippocampus needs to receive inputs which are differentiated with respect only to the areas that are their sources, but not substantive information in terms of neural firing patterns. All that is then required is for the hippocampus to compare the strength of the different inputs, ignoring any that are relatively weak (as they would not create a problem for a 'winner takes all' goal-selection mechanism). Where more than one nearly equal 'winner' remains, that is the most active input is closely matched in activity by one or more others, conflict is detected. However, given the operation of ion channels and variations in membrane potential, any instantaneous comparison could be affected as much by small recent events as by large less-recent events. For this reason, in both just-checking mode (which we are now considering) and critically in control mode (which we shall consider shortly), precise calculations within the hippocampal circuitry will be aided by temporal quantization, i.e. theta activity. From the present point of view, as we have seen, Vinogradova's CA3 comparator serves an auxiliary role, passing novel stimuli along for the particular attention of the subicular conflict detector and filtering out familiar stimuli. But the latter action can be overcome by the monoamines if the familiar stimuli are important. However, although important stimuli and novel stimuli are both passed on for further action, the type of action they provoke is different. Familiar, important stimuli are simply sent forward for checking. But novel stimuli take us back to Scenario 1: there is a mismatch between expected and actual events and the septo-hippocampal system may assume direct control over behaviour if information gathering is required. What is the result of the checking process? In the present scenario, no conflict is detected in the activated representations. The results of conflict will be discussed in the next scenario.

There must also be, here, some function for the output from CA3 to the lateral septum, but insufficient work has been conducted with selective neurotoxic lesions of these regions to tell precisely what this is.

10.4.3 **Scenario 3: conflict**

It is implicit in the argument so far that there are, in fact, four different varieties of this scenario (see Fig. 10.3D,E,H). A conflict can arise either: (1) because there is a significant stimulus event (with significance signalled by inputs using as their transmitter acetylcholine, noradrenaline, or serotonin) that is not matched by a cortical or subcortical motor signal, which would otherwise indicate that the stimulus will be appropriately dealt with; or (2) because there is a cortical model activated which is matched by no significant stimulus; or, in the special case of punishment, (3) because an expected but aversive stimulus is predicted or observed concurrently with an expected or observed appetitive stimulus; or, more generally, (4) because multiple cortical representations of incompatible goals are activated. In this latter case the conflict can involve two quite different response systems or two different stimuli which must be approached via a single response system.

As noted above, entorhinal input signals the availability of a motor program to deal with the current stimulus, or the lack of any need to respond to the stimulus. It provides a 'familiar–ignore' signal. The medial septal input provides one measure of the general level of stimulus input (somewhat different in quality from the noradrenergic and serotonergic inputs); and, in the absence of entorhinal input, has the net effect of signalling that some unexpected event has occurred. In contrast to the position taken in the first edition, this medial septal input is not held to be crucial for simple behavioural inhibition of the type required for, for example, extinction. However, since chlordiazepoxide injected into the supramammillary nucleus produces a modest impairment in water-maze performance, the theta controls that pass through the medial septal area appear to be important for the fine tuning of complex inhibitory output. It appears likely from the synergistic interactions between the ascending modulatory systems (Appendix 10) that the effects of medial septal input may be potentiated by dorsal noradrenergic bundle input (and vice versa). Further work with highly selective lesions is required to resolve these issues.

We speculate here (with the backing of some anatomical evidence) that the septo-temporal gradients of various inputs reflect a topographic mapping of response systems into the hippocampus which is similar to the topographic mapping which has been demonstrated recently for its outputs (Fig. 9.5). Thus, the ventral hippocampus 'codes' for amygdala activity, while other parts code for other response systems. This mapping may be fairly coarse (with, it has been suggested, only about four true lamellae in a rat hippocampus), but the net effect of inhibitory interneurons remains to be worked out. The longitudinal and commissural connections of the hippocampus proper would then supply the capacity for comparison of the level of activity across response programming areas as a whole, and hence determine the true level of conflict. Note that, on our present formulation, this conflict is seen as involving concurrently activated goals requiring output from distinct response programming systems. The incompatible goals elicit quite different response tendencies (e.g. approach versus avoidance). Thus, in the presence of two goals (e.g. food and safety) which elicit competing response tendencies, both approach and avoidance motor program areas will be activated concurrently (at this stage neither can take control over behaviour and their activation represents potential responding). Each sends input to the hippocampal formation, which determines by integration and comparison of the level of input: (a) that more than one response tendency (goal) is present; and (b) that activation of these tendencies is roughly equivalent. The output from the hippocampus is required most when the two tendencies are most equal. The subicular output, then, feeds back the signal of conflict to the relevant motor areas with a recursive (within and between trials) reduction of that appetitive response tendency (including seeking safety) which has the greater negative affective associations.

A second type of goal conflict is found in concurrent discriminations, delayed matching, and similar tasks. In these cases, essentially the same response (e.g. a lever press) must be made in all cases, but a choice is required between two stimuli. Here, we believe that the entorhinal cortex (and its connections with polymodal association cortex) is the key structure. In essence, the entorhinal cortex will receive two separate inputs, to either of which on its own it would normally produce a 'familiar–ignore' output, since the input indicates that an existing motor program will deal with the situation. However, like the

hippocampus, the entorhinal cortex can detect an excess of input, and hence conflict and, when it does so, will produce an output which returns directly (unlike the hippocampal output) to the input areas which generate the problem. Note that this proposed mechanism is logically prior to the 'familiar–ignore' signal which would result from receipt by the hippocampus of entorhinal output; the latter occurs whether or not there is conflict. The firing of larger numbers of entorhinal cells will therefore, from the point of view of the output to the hippocampus, result from 'excess' familiarity; but, from the point of view of the output to the neocortex, from an excess of available stimuli to which responses might be directed. Like the normal hippocampal output (assumed to increase negative bias in the associative links of the various response systems to which it is connected), the output from the entorhinal cortex would increase, in the neocortex, the values of the negative associations of the two competing stimuli and hence normally allow the appropriate stimulus to control behaviour. Even in this stimulus-oriented case, then, the output from the hippocampal formation can be viewed as supplying a form of behavioural inhibition.

We have examined the various conditions which cause the subiculum and entorhinal cortex to detect conflict. Such detection is identical with the requirement for action. What then ensues? In this context, it is important to recall that the postulated outputs of the septo-hippocampal behavioural inhibition system go well beyond simple behavioural inhibition and, at least in the case of information-gathering behaviour ('risk analysis'), can involve the release of active behaviours.

10.4.3.1 *Behavioural inhibition*

In many tasks, conflict will generate what appears to be an immediate inhibition of motor programs that are in the course of execution. This inhibition occurs at the level of planning of motor programs (Miller *et al.* 1960; Brooks 1979), not that of the execution of particular movements. Thus a CS-Pun+ (see Chapter 3 for terminology) or the odour of a cat will reduce food-gathering behaviour.

In this context, behavioural inhibition should be seen as the inhibition of only some of the concurrently activated motor programs rather than necessarily inhibition of all motor responses. For example, in complex mazes, the business of the hippocampus is not to prevent responding but to suppress alternatives to the correct response. Where some measure of responding is greatly reduced, as with food gathering, this can often be viewed not simply as inhibition of that particular program but also as selection of alternative motor programs (including freezing as such a program).

As mentioned already, there are several neurally quite distinct forms of behavioural inhibition. On the current model, many of these are mediated by the hippocampus through its interactions with a variety of motor programming structures: e.g. the cerebellum (particularly nucleus interpositus), the amygdala, the prefrontal cortex, and, probably most importantly, the cingulate cortex. In each case, the hippocampus modifies information circulating between it and these structures (this circulation occurring through a variety of routes in each case) in such a way as to increase the acuity of processing via an increase in the effect of affectively negative associations. It is this process which results in any observed behavioural inhibition. Note that from an interference

point of view, the interfering response can only be inhibited if it has relatively greater negative associations. Hence, in several tasks where the controls show the effects of interference (and *ex hypothesi* are not suppressing it), hippocampal lesions have little effect. After further training, the controls may show evidence of suppressing the interference, and it is at this point that the hippocampal-lesioned subjects will show their lack of capacity to produce this suppression (Chapter 8).

At the start of its recursive operations, the hippocampal formation receives information indicating that excessive representations are active. These could be multiple representations from one source or single representations from each of two or more sources, with the representations reflecting motor programs, stimulus relations, or some amalgam of the two (see above). This kind of ambiguity is consistent with the simple summation of input intensity across many areas performed by the supramammillary nucleus in its determination of theta frequency. What follows will be essentially the same in all cases, but is most easily understood in terms of concurrent activation of a single motor representation in each of two distinct motor-programming areas.

Let us consider environmental stimuli which concurrently activate representations of both an old-established approach response and a more recently reinforced avoidance response (i.e. an approach–avoidance conflict, Fig. 10.3E). The hippocampus (receiving concurrent input from two separate and incompatible representations) detects conflict and proceeds to circulate information between itself and both of the concurrently activated motor representations. This will have two effects. First, the initially prepotent program is executed with greater restraint (more slowly, more easily interrupted by hesitations for exploratory behaviour) because of successive interruptions from the hippocampus. Second, there will be successive alterations in the effective current activation of those representations, and hence the impact that they have on working memory. An initial, near equal, activation will thus be recursively modified until that which may have been originally marginally stronger, but has a larger number of negative associations, is sufficiently weakened to prevent the control of output; while that which was originally marginally weaker will take control. The more equal the original representations, the longer will be the time taken to resolve the conflict and the more slowly the prepotent programme will be executed. Note that while the conflict is being resolved between the two major alternatives, other behaviour such as information gathering or even displacement activity could occur and, at least in the case of information gathering, could interact with the resolution of the conflict.

Where the conflict results from concurrent activation of two incompatible stimulus representations (as in relational processing), essentially the same processes occur. Environmental stimuli activate two (or more) incompatible memory representations. The hippocampus detects this concurrent activation and recursively feeds back inhibitory signals which eliminate the marginally more negative representations in favour of the most positive (but see also below).

10.4.3.2 *The 'faulty' tag*

The stimulus representation which was in control of behaviour at the time conflict was detected has its representation altered with what in English might be a tag reading 'faulty,

needs checking'. Given the properties of long-term potentiation and long-term depression (and the phenomenon of metaplasticity; Abraham and Bear 1996), this 'faulty' tag could be a simple consequence of the processes which produce the initial behavioural inhibition, requiring no special machinery for its production. The faulty tag has two consequences.

(i) On future occasions the relevant program is executed with greater restraint, that is, its level of activation is reduced, bringing it closer to the level of the interrupting program, with the consequences outlined in the section on behavioural inhibition. This tagging could come about either because the septo-hippocampal system alters once and for all whatever stimulus representations are responsible for that motor program, or because the septo-hippocampal system itself reduces the level of activation of the tagged motor program each time it occurs. Given the other consequences of the 'faulty' tag (see below), the latter alternative is more plausible. Note that the tagging of stimulus representations of 'faulty' motor programs necessarily produces effects on that program in advance of the point at which conflict was detected. Hippocampal long-term potentiation could be the means of providing this tag, but note that the effect of the plasticity is much more like the modification of the program running in a computer than it is like the storage of data. Furthermore, the 'faulty needs checking' tag must logically be independent of the stimulus and response associations which it is qualifying.

(ii) The tagged stimulus representation is given enhanced attention by the subicular and entorhinal conflict detectors. This scrutiny of goals that have gone wrong is one of the most important functions that the theory attributes to the septo-hippocampal system. During such scrutiny the full power of the subicular conflict resolution circuits is brought into play. Stimuli in the environment, but particularly those that constitute or are closely related to the source of conflict, are subjected to an analysis that ranges over as many dimensions as possible (e.g. brightness, hue, position, size, relation to other stimuli, repetition rate, etc.). The dependencies between the animal's responses and the results of its behaviour are subjected to a similar multidimensional analysis (e.g. a turn in a maze can be classified as left-going, white-approaching, etc.). Although usually internalized, this process of analysis is closely connected to the initiation and execution of certain types of (species typical) exploratory behaviour described in the next section. Note also the similarity to the processes postulated in Cohen and Eichenbaum's (1993) relational theory, discussed in Chapter 8.

The two together—internal scrutiny and exploration—make up the process described by Kimble (1975) as hypothesis generation and testing. Their efficiency depends on the capacity of the septo-hippocampal system to control flexibly the entry of stimuli at the dentate–CA3 gate. This capacity underlies the phenomenon of 'instructed habituation', discussed in relation to Kamin's (1968) blocking effect, and its obverse 'instructed dishabituation' (see Section A9.8 in Appendix 9). By selective, sequential, opening of the dentate gate to different stimuli or stimulus attributes (often originating in the cortex) and to different response attributes (often originating subcortically), the septo-hippocampal system is able successively to examine alternative descriptions of the stimulus situation that has given rise to the conflict and to examine different response alternatives to resolve it. This process of enhanced stimulus and response analysis can be prospective or retrospective. The former case offers no particular problems. It implies

that on the occasion when conflict is detected, the stimulus complex eliciting the faulty motor program is simply tagged as such. As noted above, this can be achieved by the same signal that interrupts the ongoing behaviour. Enhanced analysis (as well as slow execution and concomitant information gathering behaviour) then commences only on the next occasion when the tagged program is initiated. The second, retrospective, case may appear more problematic, but it is more fundamental to the operation of the whole system, and particularly important in relation to anxiety. It requires that, when conflict is detected, the motor program that has just resulted in conflict and the stimuli associated with it (both those that were predicted to occur and those which actually occurred) should be 'replayed' by the loops which connect the septo-hippocampal system to the areas representing the information. This allows enhanced analysis to commence immediately, essentially re-evaluating any existing affectively negative associations.

In human terms this is close to the process of rehearsal. Rehearsal is usually taken to be verbal in form, depending on what Baddeley (1986) terms 'the articulatory loop'. This verbal loop is probably instantiated in reciprocal connections between Broca's and Wernicke's areas. Loosely speaking, disconnection of Broca's area from Wernicke's leaves a patient capable of comprehension of words and of pronunciation of words, but incapable of repetition of heard words. Thus the connection in one direction allows transfer from the comprehension of the word to its production. With the reverse connection, substituting for external auditory feedback, we have the capacity for circulation between the two areas, potentially indefinitely. There is no reason in principle, however, why rehearsal should be limited to the verbal mode or to linguistic species. Indeed, Wagner (1978) has used the concept of rehearsal with considerable success to account for a number of phenomena in animal learning. It is a central assumption of Wagner's (1978) theory that rehearsal is a more likely consequence of surprising than expected events; this is isomorphic with the hypothesis, within the present theory, that rehearsal is initiated by conflict. Among the phenomena to which Wagner applied his theory is Kamin's blocking effect. This, at the least, suggests that it is not absurd to attribute rehearsal to dumb animals. Further encouragement lies in the usual picture of rehearsal as a process that circles round a loop: as we know, there are loops aplenty in the septo-hippocampal system (e.g. Fig. 10.5).

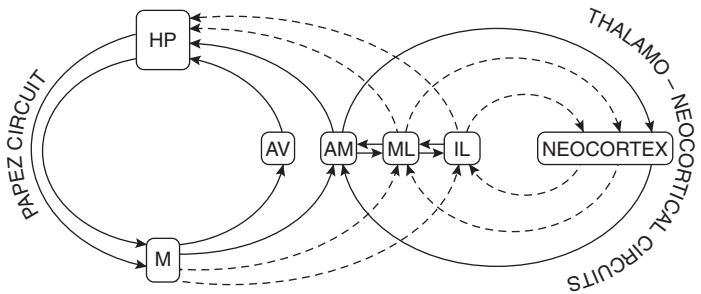


Fig. 10.5 A model proposed by Parmeggiani *et al.* (1974) to account for interactions between the Papez circuit and thalamo-neocortical circuits. AM, anteromedial nucleus of thalamus; AV, anteroventral nucleus of thalamus; HP, hippocampus; IL, intralaminar nuclei of thalamus; M, mammillary body; ML, midline nuclei of thalamus. (For further explanation, see text.)

The postulation of a rehearsal loop circling between the hippocampal formation and the representational areas with which it is connected reciprocally may offer some account of the experiments in which post-trial stimulation of the hippocampus or septal driving of theta has been observed to facilitate retention (Appendix 7). Conceivably, the electrode placements and stimulation parameters used in these experiments succeeded in activating retrospectively the process of enhanced stimulus and response analysis described above. This would have the consequence that the information whose intake preceded the stimulation would be subjected to more extensive analysis than usual and might therefore be more effectively retrieved later. 'Rumination' probably captures this process better than 'rehearsal'. Merely recycling the same information (as in verbal rehearsal) would simply strengthen whatever stimulus representations are current, both appropriate and inappropriate. We presume, in contrast, that the recursive processing of the available representations would progressively alter them in roughly the same way as in the global stereopsis example of Chapter 1 (Fig. 1.6). We have already assumed that one result of such recursion would be a more extreme separation of 'figure' from 'ground'. However, in line with the requirement to retire the faulty motor or stimulus representation as swiftly as possible and replace it with the new, more appropriate one, we must as before make the additional assumption that during such recursion, information of negative affective valence will be given more weight. The requirement for recursion lies in the fact that a simple blanket inhibition could eliminate the correct as well as the incorrect trace. Thus the amount of negative bias applied can be fine-tuned to the amount of residual conflict detected. These processes could affect consolidation during sleep (Appendix 7) as well as during waking.

In human beings, of course, we know that such rehearsal of, or rumination about, a problem can also occur in consciousness; indeed, such conscious rumination is a prominent symptom of anxiety, as considered in Chapters 11 and 13. One of us (Gray 1995) has considered elsewhere a possible extension of the present theory so as to encompass such specifically conscious concomitants of anxiety; but we do not cover this issue here.

Thus, in addition to its capacity to produce much greater strengthening or weakening of connections than would occur in a single pass, recursion (rehearsal/rumination) would produce a bias towards increased negative evaluations. This assumption will be of particular interest when we attempt to transfer the details of the theory of septo-hippocampal function to an understanding of anxiety disorders in the next chapter.

10.4.3.3 *Exploratory–investigative behaviour*

This third form of action undertaken by the septo-hippocampal system when conflict is detected is the initiation of exploration and investigation designed to gain additional information which can resolve the conflict. In a sense this is an external variation of the internal process of rumination already described. Conversely, the latter can be viewed as an exploration of internal structures so as to acquire affectively negative information which can resolve the conflict.

If one confines the discussion to spatially extended environments, this process is essentially the same as the one postulated by O'Keefe and Nadel (1978, p. 242). As we saw in Scenario 1, during initial exploration of a new environment, stimulus representations

and response–stimulus relations are analysed and stored for future use in the cortex and related areas; this is the process O'Keefe and Nadel described as the construction of a spatial map. When there is conflict, there is specific exploration of the novel element or the context that contained a now-missing element. In O'Keefe and Nadel's (1978) view a 'map' is necessary for this type of specific exploration to be initiated. This assumption is, in our view, unwarranted, as also is the assumption that animals attempt to resolve spatial conflicts differently from other kinds. On the present hypothesis, the septo-hippocampal system reacts to goal conflict in all modalities in the same way.

We can distinguish three types of situation in which the resulting exploratory behaviour can help solve an animal's problems. The first arises when the conflict is generated by a novel stimulus element that does not totally disrupt completion of the animal's current motor program. In this case, exploration of the novel element will re-establish a familiar environment by altering the current stimulus representation. This is a more limited version of the responses and reactions described in Scenario 1. The second arises when the conflict is inherent in the existing motor program, which is therefore no longer adaptive (e.g. during extinction of a previously rewarded response). In this case, exploration increases the probability of detecting a reinforcer that has been relocated in space, or which occurs under new schedule conditions. With successive failures to find any reward, the original response chain will be broken for exploration earlier and earlier, resulting, in the runway case, in the familiar pattern of extinction occurring soonest in the goal and then chaining progressively back to the run and start sections. The third arises when environmental outcomes are ambiguous and the subicular conflict resolution process (Section 10.4.3.2) poses specific questions that can be resolved by means of exploratory and investigative behaviour. For example, the output from this process may be consistent with a situation in which the walls of a T-maze are bright on occasions on which left-turning leads to food and dim on occasions on which it does not. To gain further information on this point, the subiculum would command appropriate exploratory behaviour (concerned with the brightness of the walls and the direction of the animal's turn) coupled with appropriate sensory selection. As noted above, this part of the theory is essentially the same as Kimble's (1975) proposal that the hippocampus is concerned with testing the animal's hypotheses.

We do not propose that the septo-hippocampal system itself contains the motor programs for exploratory behaviour, but rather (as for the other functions it discharges) that it modulates the use of those contained elsewhere. Thus, inhibition of prepotent responses could be all that is necessary for exploratory behaviour to be released. It is also possible that the subicular area provides direct excitation to systems that control exploratory behaviour (locomotion, eye-movements, vibrissal movements, etc.). This could occur via its projection to the cingulate cortex (Meibach and Siegel 1977). This would be in line with the proposed involvement of the cingulate in the control of a variety of innate motor programs, possibly originating in the striatum, and would be particularly consistent with the excessive checking in obsessive–compulsive disorder (Chapter 11). Excitation of systems controlling exploratory behaviour could also occur via the subicular projection to the nucleus accumbens and thence to the superior colliculus, the mesencephalic motor area, and thalamocortical sensory pathways (Gray *et al.* 1991).

Nor do we propose that the septo-hippocampal system gathers and stores the information made available by exploration. In the absence of a hippocampus, the animal learns simple discriminations as effectively as controls, e.g. in terms of trials to criterion. Thus, the power of extra-hippocampal systems to store the relevant CS+:Reward and CS–:Non-reward associations is intact. What is lost after hippocampectomy is the capacity to switch between sampling of the CS+ and the CS–. As a result, the formation of each of the positive and negative associations is carried out (as originally noticed by Kimble) in very long runs of similar choice trials. From this we conclude that in the more complex cases where hippocampal lesions appear to produce a learning deficit, the relevant simple associations can probably be formed as easily as before, but the sampling of alternatives is faulty as the result of a failure to inhibit tendencies with significant negative associations. This account is consistent with the various cases in which errorless acquisition procedures eliminate spatial navigation deficits in animals with hippocampal lesions (Appendix 8).

10.4.4 **Scenario 4: disengagement**

Conflict engages the septo-hippocampal system. Eventually it must disengage. What happens then? We consider four cases.

The first arises when the mismatch is of no fundamental importance to the animal. This is the case studied in Vinogradova's (1975) experiments on habituation. We have already considered the machinery which is brought into play in this relatively simple case. The outcome is that entorhinal input shuts the hippocampal gates by sending a 'familiar–ignore' signal. Much the same analysis applies to exploration of a novel environment. In each of these cases the role of the hippocampus is to inhibit prepotent responses and habituation of hippocampal responses reflects a loss of this inhibition. Neither habituation of the response initially elicited by the stimulus itself nor the orienting reflex is controlled by the septo-hippocampal system. Latent inhibition (see Section A9.8 in Appendix 9), can be viewed as broadly similar but with little or no initial reaction to the habituated stimulus. The partial reinforcement extinction effect (Section A9.8 in Appendix 9) is similar to latent inhibition, but where the habituated stimulus is the internal autonomic response or the memorial representation of the omission of an expected reward and processing depends on noradrenergic as well as more specific inputs. An important point to note about latent inhibition and the partial reinforcement extinction effect is that the 'familiar–ignore' signal, while no doubt affecting the hippocampus proper, has its most obvious subsequent behavioural effects through the direct connection between the entorhinal cortex and the nucleus accumbens.

The second case arises when the mismatch is important and continued performance of the plan which eventuated in it cannot solve the problem it poses for the animal. Under these conditions, the subicular comparator will ensure that the 'faulty' program is executed with increasing hesitation, that exploratory activities are undertaken, and that behaviour is varied in the attempt to light upon a more effective program. If alternative response strategies are possible (e.g. taking a different turn in a maze), the subicular comparator will have the effect of instructing them to be tried out.

Note that all of these apparently sophisticated functions of the subiculum are in reality the consequences, elicited from its target structures, of very simple subicular output. The

'faulty' tag sent by the subiculum represents, in effect, no more than a decrease in the probability of occurrence of the originally prepotent response. Where a novel stimulus is present, this inhibition will allow exploratory motor programs to take control of behaviour. Thus the subiculum most probably releases exploratory areas indirectly via the nucleus accumbens rather than controlling exploratory behaviour directly itself. Similarly, given a sufficient decrease in probability of the originally prepotent response, and in the absence of a high degree of activation of general exploration, any familiar stimuli in the environment which have some positive association will be able to control behaviour, since the originally prepotent response no longer overshadows them. Thus, the 'instruction' to try out response alternatives is a simple indirect consequence of the suppression of the original response.

If none of these various alternatives leads to success, diversification of behaviour eventually comes to an end and the animal then engages typically in some form of unlearned behaviour (e.g. grooming, sleeping). Under these circumstances the septo-hippocampal system reverts to 'just checking'.

The third case arises when the solution to the animal's problem requires new and more complex distinctions to be made between environmental stimuli, particularly under circumstances in which relational discriminations must be made. This is the kind of problem analysed by Hirsh (1974) in his contextual retrieval model of hippocampal function and by Cohen and Eichenbaum (1993) in their analysis of relational processing (Chapter 8). Like Hirsh, we suppose that the task of the septo-hippocampal system is to discriminate between contexts in which each of several competing responses is variously correct and to tag information (stored, as before, outside the septo-hippocampal system) with an appropriate 'contextual label'. Absence of this function, as in an animal with hippocampal damage, would lead to excessive susceptibility to interference from competing responses, as is observed. Like Cohen and Eichenbaum, we see one critical property of such stimuli, and hence of the required labelling, to be their relational character. However, we take the diametrically opposite position to them as to the function of the septo-hippocampal system, and this gives us a simpler view, also, of Hirsh's tags. In our view, as discussed above, the business of the hippocampus is to eliminate incorrect connections which will normally result from the simple associative rules instantiated in long-term potentiation. It does this by suppressing the conflicting behaviour during the early phases of conditioning and by adding a 'faulty' tag to the association which gives rise to that behaviour. Ultimately, long-term potentiation of the correct (e.g. A–B–C) relation and long-term depression of the incorrect (e.g. A–C) link will lead to permanent changes so that A–C no longer tends to elicit any response. At this point only one response tendency (appropriate to A–B–C) is elicited, there is no conflict, and the septo-hippocampal system reverts to 'just checking'. Thus, again, apparently complex results of hippocampal output (the formation of relations; the addition of contextual tags) are simply a consequence of the way extra-hippocampal systems react to receipt of the interrupting output from the septo-hippocampal system (particularly its increase in the weighting of affectively negative associations of goals) and do not assume any complex cognitive processing by the latter system itself. 'Relational' processing may in this way be viewed as a simple extension of normal associative processes; these create the multiple relations implicit in the situation, from which interaction with

the hippocampus carves away adverse associations until only the behaviourally effective one remains.

The fourth and final case to be considered is that in which the septo-hippocampal system can never disengage without disrupting performance. In any of the cases in which the hippocampus can be viewed as promoting the formation and storage of memories of appropriate associations (by weakening competing, inappropriate associations), the septo-hippocampal system should ultimately be able to disengage and, in at least some cases, suitable training methods may be able to bypass the requirement for the system to play a part even during acquisition. But some deficits produced by septo-hippocampal damage are apparently absolute, that is, they show little if any sign of recovery even over prolonged periods of testing (Olton *et al.* 1979a). These tasks have in common the use of an ever-changing list from which the animal must choose the correct alternative, itself ever-changing. The best examples of tasks of this kind come from Olton's work with the radial-arm maze (Olton *et al.* 1978b; Olton *et al.* 1979a) and Gaffan's (1977a,b) experiments on recognition memory (see Chapter 8). The critical difference between tasks of this fourth kind and those susceptible to analysis in terms of contextual labelling is probably that now even the context cannot be specified in absolute terms. Compare a conditional discrimination with Gaffan's (1977a) experiment in which monkeys had to respond to the second appearance of each of a list of 25 pictures (Chapter 8). In a conditional discrimination, it is possible to label the correct responses once and for all: for example, go left if the walls of the maze are white, go right if they are black. But in Gaffan's experiment only the general rule (respond always to the second appearance of a slide) can be stored permanently, and this is useless as a guide to action unless the particular list of pictures to be responded to is also constructed afresh each time they occur.

This analysis has much in common with Olton's (1978a) in terms of working memory:

> The system begins with a sensory input that enters into a temporary register. A comparison process attempts to match the contents of the temporary register with each of the items in working memory. A match indicates that the choice in question (of one of the arms of the maze) has already been made and should not be repeated; information in the temporary register is deleted, a decision made as to whether to reset working memory, a search is initiated for a new choice, and new sensory input is obtained. A failure to match the contents of the temporary register and some item in working memory indicates the choice in question has not been made previously and ought to be made now. Running down the arm produces reward. The information defining the choice which was in the temporary register is stored in working memory so that the choice will not be repeated, the temporary register is cleared, the reset decision is made, a search for another choice is initiated, and new sensory input is obtained. (Olton 1978a, pp. 363–364.)

In a minor departure from the first edition, we accept the widespread view that working memory in its purest form depends on areas such as prefrontal cortex (Chapter 6; Appendix 3) rather than the septo-hippocampal system. With relatively novel and highly discriminable stimuli in a simple delayed match-to-sample, the recent association of the target stimulus with its secondarily rewarding sample context will make it clearly distinguishable from the comparison stimulus. The two stimuli will have quite distinct values and only the target will be treated as a goal. However, as discussed in greater detail in

Chapter 8, previous association of a stimulus with reward, especially if combined with some factor which makes stimuli less discriminable within the memory domain (e.g. the use of simple as opposed to complex stimuli), will allow secondary rewarding value to accrue to the comparison stimulus either by prior conditioning or by generalization. The approach and avoidance tendencies inherent in the choice will then be more evenly balanced, there will be conflict, and the septo-hippocampal system will be required to boost negative associations and hence prevent the incorrect choice from being made. In tasks of this kind, the problem of interference becomes worse rather than better with experience of the stimuli, and so the hippocampus can never disengage once it has become involved in the first place.

A final word should be said here about passive avoidance. This might be thought to be the quintessential paradigm in which a behavioural inhibition system should be involved. Like the tasks we have just considered, therefore, it might be expected always to require a hippocampus whatever the variant of the task and whatever the length of training. This is not, however, the case. In all our examples so far it has been the conflict between two equally balanced tendencies which has required hippocampal resolution. While this will often be the situation during the acquisition of some passive avoidance tasks, it is not a logically necessary feature of the paradigm, even during acquisition. Consistent with this view, our analysis of the data in Appendix 8 shows that the size of effect of septo-hippocampal lesions in passive avoidance tasks is directly related to the extent to which approach tendencies (accidentally or intentionally elicited by the procedure) have been as strongly activated as the avoidance tendencies. Here again, conflict is more fundamental than behavioural inhibition as such; but behavioural inhibition is the means the septo-hippocampal system uses to eliminate conflict.

It is probably as well, here, to emphasize once more a feature of our analysis which will make it unpalatable to those who like sweeping paradigmatic generalizations and theories which can be reduced to a single sentence or adjective.[4] In all cases, 'conflict' is the result of the perception *by the subject* of two or more incompatible goals. These goals are defined by the subject and not always recognizable to the experimenter. As a result, conflict will not be absolutely predictable from the official (learning theory) paradigm. This is why there are virtually no cases in our original review of the lesion literature (Gray and McNaughton 1983) where there are completely consistent results, and why exceptions involve special features (e.g. the use of olfactory stimuli in simultaneous and successive discriminations) which can be understood post hoc, but are not readily predictable in advance.

As we note in Chapter 8 when discussing similar problems with the concept of interference, the present version of the theory cannot specify exactly where in a range of parametric variations a paradigm will become hippocampal sensitive. The theory is, nonetheless, readily tested in that it makes the unequivocal prediction that, whatever the official nature of a paradigm (e.g. procedural rather than declarative memory), it will be possible to create a version sensitive to septo-hippocampal damage by any one of a variety of manipulations which introduce goal competition. This seems a reasonable claim given

4. Our thanks go to Ray Kesner for pointing out the necessity to make this clear (and for the many hours he spent constructing thought experiments to challenge our theory).

the many cases in which we know it already to be true. We are equally certain that any paradigm can be produced in a form which is insensitive to septo-hippocampal damage if we first determine the nature of, and then remove, inherent sources of goal competition. This is, of course, difficult to achieve in practice, since competition may come from some unobvious species-typical response (as in the case of Cohen and Eichenbaum's analysis of simultaneous olfactory discrimination; Chapter 8). Note also that, as with our discussion of interference, it is not the presence of the competition, alone, which will render the task hippocampal sensitive. It is also necessary that control animals should be capable of suppressing the competing response (see Chapter 8 for a relevant experimental demonstration).

10.5 THE THETA RHYTHM

There remains one major phenomenon to incorporate into the theory: the enigmatic theta rhythm. We have retained a weak form of the common view (e.g. Vinogradova 1975; O'Keefe and Nadel 1978) that theta is involved in timing, quantizing, and indexing information; or, more strictly, that when theta is observed in the EEG it reflects the phase locking of large numbers of hippocampal cells, and this phase locking is important if these functions are to be achieved. We consider the various strong forms of this hypothesis, first, and give our reasons for rejecting them.

In O'Keefe and Nadel's (1978, p. 220) model, theta plays a critical role in determining which neurons store the information that constitutes a spatial map and subsequently in retrieving this information when it is needed. It is certainly the case that information arriving at a particular point of the theta cycle can preferentially undergo long-term potentiation (Appendix 5). A dependence of this kind is likely in all areas where phasic bursting of cells is found. However, the more specifically the control of theta is impaired, the less the resultant manipulation reproduces the entire septo-hippocampal syndrome; in particular, a specific reduction in theta frequency (which might be expected to grossly impair any indexing function) has only modest effects on spatial navigation (Pan and McNaughton 1997). It appears, then, that loss of theta degrades hippocampal processing but does not eliminate it.

An equally interesting view is that different theta frequencies may determine which loops in the system achieve priority in recursive processing, by matching the period of the theta wave to the time taken for information to circulate in a particular loop. This possibility is supported by the observations of Parmeggiani *et al.* (1974), who investigated the passage of neural messages around the loop comprising the subiculum, mammillary body, thalamic nuclei, cingulate cortex, and back to the subiculum as a function of frequency of stimulation of the hippocampus in anaesthetized cats. Multiple-unit responses in the mammillary bodies and the anteroventral nucleus of the thalamus were a function of stimulus intensity only; but the response recorded in the anteromedial nucleus of the thalamus showed a sharp resonance at about 7 Hz, response magnitude falling rapidly as stimulation frequency rose above this value. Parmeggiani *et al.* (1974) interpret these results as indicating that the anteromedial nucleus acts as a gate, circulating information of hippocampal origin to the neocortex (to which it projects) only at

frequencies below 7 Hz in the cat (Fig. 10.5). A mechanism such as this might play a role in selecting information of neocortical origin which is then fed back to the septo-hippocampal system via the entorhinal area. Indeed, in the first edition, based on Parmeggiani *et al.*'s (1974) results, we suggested that 'different frequencies of theta might determine, not only the rate at which the septo-hippocampal system processes information, but also *which* information is processed' (Gray 1982, p. 293).

An extensive and formal version of this hypothesis has been put forward more recently by Miller (1991), who proposes that specific theta frequencies act to select contexts for the storage and retrieval of information:

> The total conduction delay time round each one of these patterns of loops is envisaged to correspond to the theta period. Such resonance between hippocampus and cortex is envisaged to have an important functional role in the registration and retrieval of information in the cortex. Each pattern of resonant loops will raise the activation of specific collections of cells which are widely dispersed across the cortical mantle. (Miller 1991, p. 159.)

Thus resonance, during storage of information, will cause stimulus information to be stored in a particular (frequency specific) context. Similarly, the same resonance during retrieval will reinstate the original context and hence enhance retrieval. 'The hippocampus thus acts in some ways as an "index" to memories, the cortex being, as it were, the "book" to which this index refers' (Miller 1991).

This is a very attractive notion, particularly as information storage is proposed to occur in the cortex rather than the hippocampus. However, it has two flaws which cause us now to reject, not just its specifics, but also the principle of circuit selection by frequency encoding. First, although information coding is postulated to occur in the cortex, its retrieval is as dependent on the frequency of theta as its storage. This makes the theory as much a memory theory of the hippocampus as the others we have considered. As such, it cannot handle the data on interference and intrusion errors or intact retrograde memory after hippocampal damage. Second, the relationship of theta to behaviour and the accuracy of tuning of different loops make it virtually impossible that the model could work. Theta frequency varies with an animal's level of arousal rather than with the context in which it finds itself. High-frequency theta can be obtained when the animal is freezing hard or running hard. Low-frequency theta can be obtained when the animal is still or producing modest movements. Theta frequency is highest at the initiation of movement and then settles down to a frequency which is essentially constant during the remainder of a trajectory. In none of these respects can frequency provide an unambiguous or reliable context, and in probably no case will theta frequency be the same during storage and retrieval of a particular piece of information. For the same reasons, theta frequency cannot be used to select circuits (and particularly structures) of an assigned significance.

One exception to this negative conclusion should be noted. We have rejected the idea that theta frequency can encode specific contexts and hence select particular classes of information. However, in our discussion of septal driving of theta, we suggest (Appendix 7) that frequencies in the region of 7.7 and 6.9 Hz could produce selective results in relation to circuits that are sensitive to locus coeruleus and raphe input respectively. In

these cases the selectivity is with respect to inputs which we have already concluded are modulatory and related to (somewhat different) aspects of attention, arousal, stimulus intensity, or motor activation. Furthermore, whatever one's label for this underlying source (or sources) of activation of the noradrenergic and serotonergic systems, such activation is likely to show a high correlation with theta frequency. Thus, any tuning that is present is likely to be specific to the current emotional state of the animal, but orthogonal to the specific information being processed (whether this is contextual or otherwise).

We have, then, rejected any overall timing function for theta, but there remains the putative quantization function. Parmeggiani *et al.* (1971) have shown that the conduction time round the subiculum–mammillary body–anteroventral thalamus–cingulate–subiculum loop is about 50–60 ms in the curarized cat. Miller (1991), in the context of his resonance theory, has argued that many of the subiculo-cortical pathways have loop conduction times of the order of 100–200 ms. There is clearly, then, the possibility of considerable drift if recursive calculations are being made which involve several of these loops at once. On the currently available information, the function that we would, therefore, ascribe to theta activity (that is the phase locking of firing of cells, even if this does not produce a noticeable theta rhythm) is the exact opposite of the 'circuit selection' function. We believe that theta activity maintains the discreteness of individual cycles of recursive calculation, and thus ensures that the inputs from all the different loops are dealt with by the various hippocampal comparators at essentially the same time on a consistent basis.

One example of the type of operation we have in mind is provided by Komisaruk (1970) and Macrides (1975). They showed that theta and vibrissal movements sometimes become phase locked for short periods and that it is movements of the vibrissae which are entrained to theta. At times like this, we suggest, phase locking ensures that successive chunks of information reach the hippocampal system from the vibrissae or via olfactory channels closely coupled to the vibrissae, optimally packaged for analysis. Given the old view of the hippocampus as the rhinencephalon, it may even be that the frequency of sniffing is the evolutionary origin of the requirement for phase locking of the system, this original mechanism now being vestigial. Equally, the capacity of important stimuli to reset the phase of theta (Givens 1996b) makes sense in terms of the fact that the inhibition of hippocampal circuits represented by the resetting will effectively clear the system with respect to the previous calculations and make it ready for processing of the newly significant item.

There are some important consequences of this synchronizing view of theta. First, there is no need for any great rhythmicity in cortical as opposed to septo-hippocampal cells, since resonance as such would not be a critical feature of cortical operation (although of course some rhythmicity would be expected simply as a consequence of the rhythmicity of hippocampal output to those structures). Second, removal of theta would not eliminate all hippocampal function. Theta would be required particularly for those calculations in which difficult (high iteration number) separations of conflicting alternatives are undertaken, but would be progressively less necessary as the conflict becomes easier to resolve. Third, theta would be expected in both 'just checking' and 'control' modes, as it would quantize the receipt and comparison by the hippocampus of information; this should occur independently of whether the hippocampus detects conflict in that

information or not. However, theta would be functionally critical only in relation to control mode, that is when recursive processing is occurring and functionally significant output is being produced.

10.6 CONCLUSION

This completes the presentation of the theory, the circuitry of which is summarized in Fig. 10.6. It can account for many of the data. Since it is largely based on these data, this is not surprising. In the first edition we said that 'only future experimental test will tell whether it has any real merit'; this still—perhaps surprisingly, nearly 20 years later—remains the case. Although we have made some alterations of emphasis and of particular details of mechanism in the current edition, the basic theory is unchanged. We see it as a strength that it encompasses the early data which recent memory theories largely ignore, yet has withstood the data obtained since the first edition, on which the memory theories are largely based.

How much is the theory our own? It has borrowed much from others. In the literature we have covered, it would have been something of a disaster if none of the insights gleaned by earlier workers were worth preserving after the massive investment of research into the septo-hippocampal system, and related defence and frontal systems, over the last half-century. We have, however, been selective in our borrowings and have felt free to make alterations of which the original authors might well not approve. In all cases the basis for the selection and for the alteration has been the attempt to produce a consistent theory that encompasses all the data at all levels of analysis. Inevitably, there will be those who dislike the way we have put our borrowings together and who will find the result a patchwork quilt, no stronger than its patches and the weaker for the seaming. On the contrary, we see the overall theory as stronger than the individual elements it has striven to put together: a subtly blended cocktail, perhaps, rather than a patchwork.

Central to the distillation is the notion that the septo-hippocampal system serves as a conflict detector. For this it contains a comparator between expected and actual stimulation, not unlike those proposed in different ways by Vinogradova (1975), O'Keefe and Nadel (1978), and Olton (1978a). In our formulation, this comparator function is more general, more goal-oriented, and recurs at different points in the hippocampal circuitry; but nonetheless it retains the fundamental properties common to these earlier positions. From Kimble (1975) we have taken the view that the septo-hippocampal system generates and tests hypotheses under conditions of environmental uncertainty. In this edition we attribute the actual generation and testing of hypotheses to the goal-processing systems which are unmasked when a prepotent goal receives a 'faulty, needs checking' signal from the hippocampus. But the interaction of these goal-processing systems with the septo-hippocampal system is critical for this function. So our position is virtually identical to Kimble's for the data to which his view can be applied. From Hirsh (1974) we have taken the idea of contextual labelling, and from Winocur (1981) the idea that animals with hippocampal lesions can be overly dependent on contextual cues; and we have produced an amalgam of these two positions. From Butters and

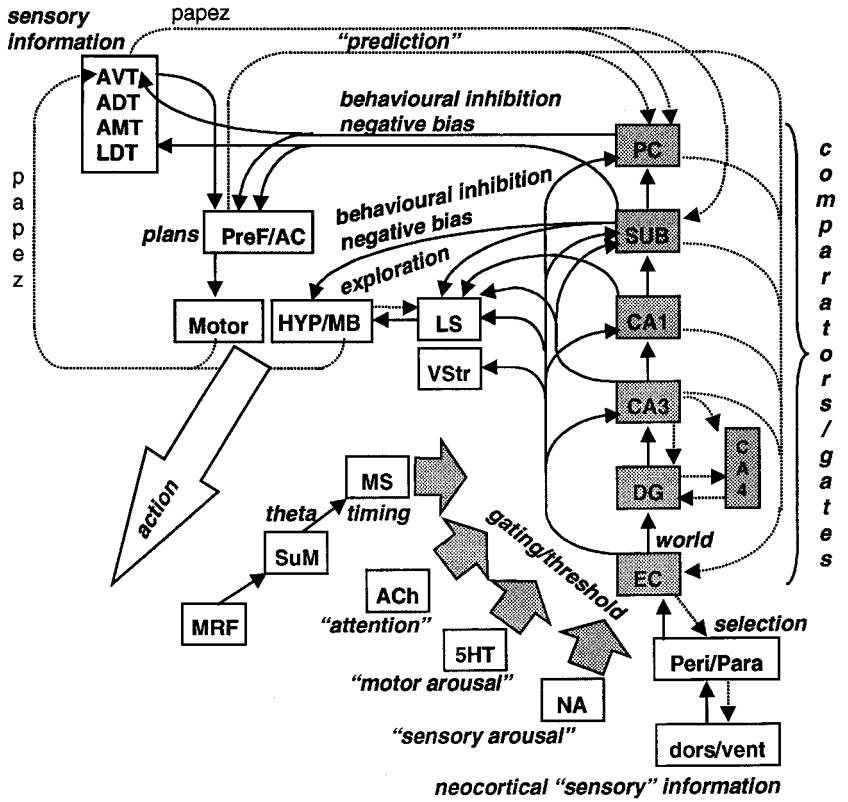


Fig. 10.6 A summary of the septo-hippocampal circuitry subserving the behavioural inhibition and attentional outputs of the behavioural inhibition system. The circuits presumed to deal with perception and action have been simplified from Fig. 9.4 and those controlling theta rhythm have been simplified from Fig. 10.2. For the detailed workings of the system see the earlier parts of Chapter 10. The overall architecture follows that of Fig. 1.4. The arousal output of the behavioural inhibition system is held to be mediated independently by the amygdala and has been omitted. 5HT, 5-hydroxytryptamine; ACh, acetylcholine; HYP, hypothalamus; Motor, motor systems; MRF, midbrain reticular formation; MS, medial septum; NA, noradrenaline; SuM, medial supramammillang nucleus. For other abbreviations, see Fig. 9.4.

Cermak (1975) we have taken the idea that multidimensional stimulus analysis is important, and that this is why the linguistic channel appears particularly liable to the effects of hippocampal damage in human beings. From Cohen and Eichenbaum (1993) we have taken the idea that relational analysis is important.

But, while the theory presented here owes much to these (in some cases relatively recent) concepts, it unashamedly harks back to and incorporates the earlier view of the septo-hippocampal system as playing an essential role in the inhibition of maladaptive behaviour (McCleary 1966; Douglas 1967). Many workers in this field, particularly with the recent efflorescence of memory theories, appear to believe that one has to choose between a model in which the septo-hippocampal system analyses stimuli and one in

which it acts on responses (see, for example, the discussion in Elliott and Whelan 1978, p. 306; and the disavowal of non-mnemonic functions of the hippocampus in Eichenbaum *et al*. 1994). But this choice is not only unnecessary, it distorts the problem in two different ways. First, on the view developed here, the information-processing and behaviour-inhibitory roles of the septo-hippocampal system are, functionally speaking, two sides of the same coin. These roles are, of course, discharged by different particular neurons, since they reflect the input and output sides of the hippocampal circuitry, respectively. To separate these functions and ask which of them is carried out by the hippocampus is to forget that we are dealing with a level of the nervous system where perception (detection of a specific class of input) and action (production of a particular class of output) are essentially synonymous. Our frequent emphasis on this equation is again borrowed, this time from MacKay (1987). Second, we have also emphasized, particularly in this second edition, that the septo-hippocampal system deals with goals and plans. A goal has both stimulus and response properties. To attempt to remove the stimulus properties from a goal will destroy its qualities just as assuredly as attempting to remove the implied action tendency (Fig. 1.7).

In our treatment of external structures related to the hippocampus we have borrowed wholesale from recent formulations. Our view of the defence system is an amalgam of the views of Graeff, LeDoux, and Davis. Our view of the organization of frontal cortex is based largely on Pandya and Barbas. Our view of working memory is a blend of Baddeley, Goldman-Rakic, and Petrides; and our view of the basic nature and different types of memory owes much to Fuster. Our view of cingulate cortex combines the views of Vogt and Rapoport. Our view of the eye-blink conditioning circuit is essentially that of Thompson. In all these cases we have made only minor alterations to make the different authors' views fit each other, and occasionally to bring specific details into line with our overall theory. Thus, however idiosyncratic our views of the hippocampus itself, our views of the structures to which it is connected are largely those of the current literature.

Since the theory is a blend, it is, we hope, unnecessary to provide a detailed analysis of all the findings summarized in the other chapters and appendices to see how well it can account for them. Some of the findings have been used explicitly to develop the theory and to produce its recent modifications. Others are discussed fully elsewhere in the book in terms of the particular theoretical ingredients to which they pertain. The bulk has contributed to the interim conclusions of the various chapters; and it is these conclusions which have supplied the fundamental building blocks of the theory. An important feature of the theory, however, is that its integration provides strong support for the conclusions arrived at in the individual chapters, especially since these, within a particular chapter, often hinged on only a few results. The overall theory, then, often accommodates recalcitrant data better than do the individual hypotheses that tie it together.

This, then, is our theory of septo-hippocampal function. Is it also a theory of anxiety? This step is not yet warranted; it will be addressed in the next chapter.

11 Symptoms and syndromes of anxiety

In this chapter, we take the detailed knowledge we have obtained of the actions of the anxiolytic drugs, of the organization of the hierarchical defence system, and of the functions of the septo-hippocampal system and consider the implications for the diagnosis of clinical anxiety and anxiety-related disorders. In the process we must integrate these different bodies of data into a coherent neuropsychology of anxiety and related states such as panic. In our attempt to do this, we adopt here a 'neurologizing' stance, making best guesses as to the structures that most intimately underlie each particular condition. This stance lures us into 'splitting' apart rather than 'lumping' together the various phenomena we address. Later, however, when we deal with personality and treatment in the final chapters, that strategy will undergo a significant correction.

The neurology of the defence system is critical for our understanding of the clinical situation; it is summarized by the hierarchical model of Fig. 11.1. Sensory stimuli of all modalities are held continuously to bombard the systems responsible for active defence (as well as all other systems concerned with goal-directed behaviour), through both 'quick and dirty' and 'slow and sophisticated' routes (LeDoux 1994; Chapter 6). In the latter case the sensory input results also in the activation of memory traces which provide additional input to goal-oriented systems. Output from the defence system is also hierarchically organized (Graeff 1994; Chapter 6). The lowest levels of the system produce the fastest and simplest responses. As one goes to higher and higher levels the speed of response production decreases, but the complexity of responding increases, with the most complex avoidance responses originating in the prefrontal and anterior cingulate cortex.

Concurrent activation of different levels of the defence system (undirected escape in the periaqueductal grey, directed avoidance in the amygdala) are resolved by its hierarchical organization. The control of behaviour by the lower levels is inhibited when there is a greater level of activity in the higher levels of the system. But, also important for an understanding of the clinical phenomena, too great an activation of the lower levels can allow it to escape this inhibition. Following Graeff, we have argued (Chapter 6; Appendices 2 and 10), that at least part of this control results from the opposite effects on the periaqueductal grey and amygdala, respectively, of serotonin (5-hydroxytryptamine) release. In a sense, treatments which promote the release of serotonin inform the defence system that a problem is soluble (or, more generally that a situation can be coped with). Note that in the context of our theory of the septo-hippocampal system (Chapter 10), the interaction between levels of the defence system does not involve a conflict between goals, does not involve recursive interactions, and requires only simple one-way inhibition, the level of which does not have to be adjusted according to circumstances.

In parallel with the active defence system is the passive defence system. This can be equated with the behavioural inhibition system. The defining characteristic of the behavioural inhibition system is that it inhibits prepotent behaviour, and so allows resolution of conflict between concurrent incompatible goals. This inhibition extends to future

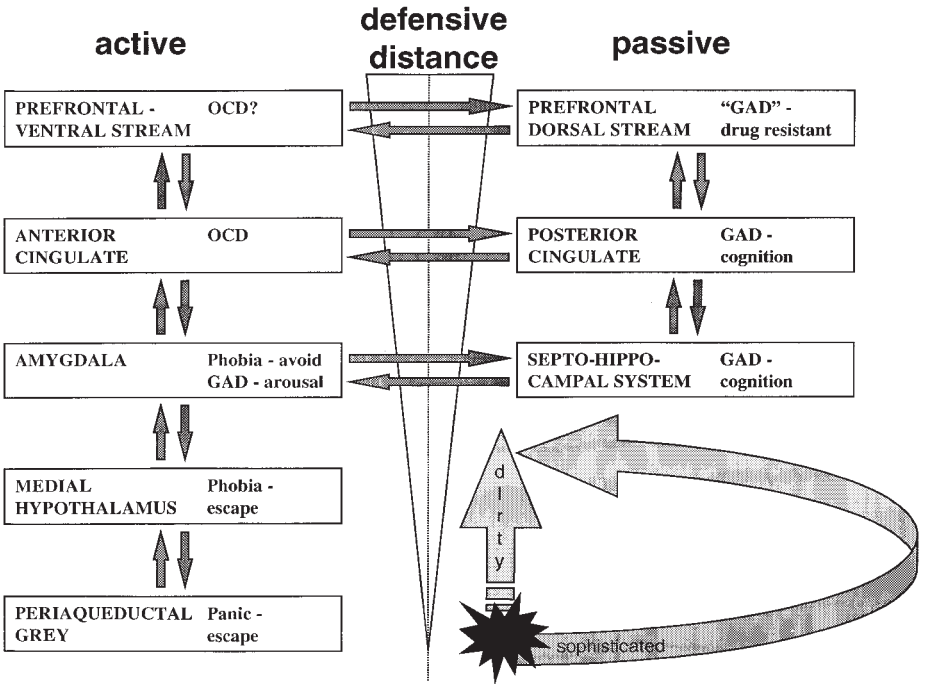


Fig. 11.1 The role of the hierarchical defence system in anxiety-related disorders. This shows the basic neural architecture of the hierarchical defence system (Fig. 1.8), extended to include prefrontal cortex. All parts of the system receive both fast poorly-digested (dirty) and slow well-digested (sophisticated) sensory information. The lowest level of the system is held to deal with the most basic response to threat—panic. Progressively higher levels deal with progressively more anticipatory reactions (defensive distance) and consist of two separate streams concerned with active avoidance and passive avoidance, respectively. While activity in the system should always be seen as distributed across several parts simultaneously and while symptoms cannot be equated with disorders, we have assigned the control of particular symptoms and/or disorders predominantly to specific areas. For further details, see text. GAD, anxiolytic-sensitive generalized anxiety disorder; "GAD", anxiolytic-insensitive generalized anxiety disorder; OCD, obsessive–compulsive disorder.

occurrences of originally prepotent behaviour through the tagging, produced by the septo-hippocampal system, of prepotent goals with increased negative associations. Equally important characteristics of the behavioural inhibition system are that it increases arousal (mediated, along with activation of the autonomic nervous system, via the amygdala) and attention, particularly to negative reinforcing stimuli. Major negative affective consequences of activation of the behavioural inhibition system are produced largely by output from the septo-hippocampal system to the amygdala, this having the function of increasing the relative strength of avoidance responses programmed via the latter structure. The septo-hippocampal system can also inhibit all areas which process appetitively motivated goal-directed behaviour, effectively through an increase in negative cognitive bias.

Anxiolytic drugs, as a class, impair the functions of the septo-hippocampal system. The precise nature of this effect is critical for the clinical application of our theory, and

is considered in detail below. The effects of anxiolytics on the subcortical control of theta activity (Appendix 5) seem sufficient to account for the bulk of their relevant behavioural effects (those which are like the effects of damage to the septo-hippocampal system). In addition, the anxiolytic drugs have a direct action on the amygdala. This accounts, critically, for their effects on conflict-induced arousal, effects that are not mediated by the septo-hippocampal system. The action of the anxiolytics on the amygdala is, however, limited. Thus, as reviewed in Appendix 2, although direct benzodiazepine infusions into the amygdala (like amygdalar lesions) impair behaviour in a number of tests of anxiety, such as passive avoidance, the similar effects of systemic benzodiazepines are not necessarily blocked by amygdalar lesions; and, conversely, there are effects of amygdalar lesions (e.g. on active avoidance) that are not produced by intra-amygdalar benzodiazepines. We interpret this pattern of data as showing that the action of anxiolytics on the amygdala blocks an input that biases its activities towards the production of passive rather than active avoidance. We propose, further, that this input derives from the septo-hippocampal system when the latter detects conflict, its role under these conditions being to amplify both current and future passive avoidance outputs from the amygdala. This set of actions of the anxiolytic drugs on the septo-hippocampal system and the amygdala, taken together, is able largely to account for their clinical anxiolytic action, although minor contributions from the hypothalamus (in addition to that due to changes in theta rhythm) and periaqueductal grey are possible.

This theory, then, attributes to the septo-hippocampal system the capacity to increase the valence of affectively negative stimuli (thus, in many cases, increasing perceived threat). It follows that hyperactivity of the septo-hippocampal system should produce a fundamentally cognitive dysfunction (excessive negative bias; see below), the most obvious consequences of which will be affective (excessive anxiety), because this bias will increase the immediate perception and hence subsequent storage of threatening associations under conditions of conflict. It is this hyperactivity which produces generalized anxiety disorder, and on which the anxiolytic drugs exert their clinically most important effects. In contrast to this 'core' anxiety and lying, as it were, outside its bounds, we attribute phobia (in the sense of simple active avoidance) principally to the amygdala and medial hypothalamus; and panic, to the periaqueductal grey.

There is in addition, however, an anxiolytic-insensitive form of anxiety, in which clinical benefits are seen, not with administration of these drugs, but after surgical disruption of descending projections from the cingulate or prefrontal cortex (see Section 6.13). In the 1970s patients who showed these benefits were described, pre-operatively, as having predominantly obsessional symptoms, widely generalized phobic behaviour, or anxiety states (Marks *et al.* 1966; Tan *et al.* 1971; Powell 1979). More recently, clinical practice has concentrated on obsessive–compulsive disorder and major depression (in each case only if resistant to other forms of treatment) as the primary indications for such 'psychosurgery' (Scottish Office 1996). It would make our analysis of the neural basis of anxiety much easier if the only anxiety-related disorder suitable for psychosurgery were indeed obsessive–compulsive disorder. We have in any case other reasons (see below) for distinguishing this condition from the core anxiety state. Thus, if obsessions and compulsions were the only anxiety-related symptoms responsive to psychosurgery but not to anxiolytic treatment, we could rest easy with the criterion of sensitivity to the latter as picking out 'true'

anxiety. However, a review of the psychosurgery literature by Waziri (1990) does not justify this conclusion. Waziri's Table 1 shows that out of 183 patients 'with anxiety' operated in 10 studies, 49% were either 'symptom-free' (49 patients) or had only 'minor symptoms' (40 patients) at follow-ups ranging from 8 months to 20 years. The equivalent figures for patients 'with obsessions and compulsions' (Waziri's Table 2) were, out of 232 patients operated in 11 studies, 54% either symptom-free (73 patients) or with minor symptoms (53 patients). The slight advantage for the latter group of patients does not justify the recommendation in the report of the Scottish Office (1996, para. 67) that obsessive–compulsive disorder should be the only indication among the anxiety disorders for surgical treatment. This recommendation is also not justified by the studies (overlapping only partially with those in Waziri 1990) reviewed in Annex D of the report itself, where again other anxiety disorders appear to benefit from psychosurgery to a degree not greatly different from that shown by obsessive–compulsive disorder.

Clinically, then, the recommendation of the Scottish Office may well be correct if other factors, such as the availability of alternative treatments, are taken into account; but scientifically, besides obsessive–compulsive disorder, there clearly are other types of anxiety disorder resistant to pharmacological treatment but which benefit from disruption to pathways descending from the cingulate and frontal cortices. There is other evidence (see Appendix 3) indicating the cingulate cortex as important in the production of the symptoms of obsessive–compulsive disorder. As to the prefrontal cortex, the role played by this region is likely to consist in bringing to bear outputs from neocortical language systems onto the activities of lower structures. In these ways, then, the effects on anxiety of psychosurgery can be integrated with other data and concepts. Nonetheless, these effects force upon us the complexity of accounting for forms of anxiety that are both sensitive and insensitive, respectively, to anxiolytic drug treatment.

Despite this added complexity, we still have a basically simple scheme that can be applied to both the symptoms and syndromes of anxiety-related disorders. However, the application is not straightforward. First, there are extensive reciprocal connections between the different components of the defence system, and many of these involve activation (or release) rather than inhibition. This implies that they should all interact. Furthermore, an important function of the ascending modulatory aminergic systems and of stress hormones is to modulate these interactions; and these systems are themselves subject to long-term change in response to stress (Toates 1995). Second, it is not a stimulus as such which gives rise to anxiety, but rather the interpretation placed on the occurrence of that stimulus by the individual, animal as well as human (Chapter 3). The inputs to the defence system, therefore, are subject to cognitive elaboration and to processes such as observational learning. It should not be assumed, however, that complex cognitions necessarily produce complex responses. For example, area 23 of the cingulate cortex, which receives input from the highest levels of unimodal and polymodal sensory cortex, not only sends output to prefrontal cortex but also has a direct projection to the periaqueductal grey. Through this route cognitive modification of the symptomatic and behavioural outputs that constitute panic can be achieved, but these outputs themselves remain relatively stereotyped.

In applying this theory to the rich phenomena of the clinic, we need first to clarify the nature of anxiolytic action. From this we shall derive a view of generalized anxiety

disorder as, in at least some cases, reflecting a cognitive dysfunction. We shall also derive a typology of anxiety from the neurology of the defence system. We then apply the resultant theory to a number of the current categories of anxiety-related disorder.

11.1 COGNITIVE EFFECTS OF ANXIOLYTIC DRUGS

Appendix 8 (see also Chapters 4 and 6) documents an extensive similarity between the behavioural effects of septo-hippocampal damage and those of the anti-anxiety drugs. However, the two literatures have proceeded independently and there are many tasks (usually those of most theoretical significance) for which we have data on the effects of lesions but none on those of drugs, and vice versa. In particular, hippocampal damage in human beings is characterized by amnesia. An important question is, therefore: how far do moderate, anxiolytic, doses of the drugs affect memory tasks?

At the time of the first edition, Sahgal and Iversen (1978) had shown that chlordiazepoxide impairs the ability of pigeons to make a same-difference judgement. There were few other data which linked anxiolytic drug action with the clearest 'non-emotional' tests of hippocampal dysfunction. We speculated then that

'the septo-hippocampal syndrome may yet turn out to be wider than the one created by anti-anxiety medication. Indeed, if it has anything to do with amnesia, it must be wider, since this is at least not an obvious feature of the behaviour of patients treated with anti-anxiety drugs. But note that when used in large doses to induce relaxation as a part of anaesthetic procedures, the benzodiazepines have been reported several times to cause anterograde amnesia for material memorized at that time' (Gray 1982, p. 298).

Indeed, the high doses of benzodiazepines used as pre-medication for a variety of procedures do produce an apparent amnesia. However, this could be attributed to their sedative effects or some other non-specific factor such as changes in vigilance (McGaughy and Sarter 1995). Several studies with more modest doses have nonetheless indicated effects on memory, and these are of a type which has led to the inference that 'the benzodiazepines might simulate, or model, the form of memory dysfunction . . . that is expressed in the typical amnesic patient. . . . [Further,] recent studies . . . suggest that the sedative and memory impairing effects of benzodiazepines can be dissociated from one another' (Danion *et al.* 1993, p. 373, see for brief review; also Curran 1991). 'Memorial' effects appear to be obtained more readily with lorazepam than with diazepam or chlordiazepoxide. However, in some circumstances at least, diazepam too can have effects equivalent to those of lorazepam. These effects appear to be on the capacity to acquire information (Vidailhet *et al.* 1994).

In animals, as discussed in Chapter 4, anxiolytics, including buspirone, produce hippocampal-like impairments in the acquisition (but not performance) of the water-maze task (McNaughton and Morris 1987, 1992). We have also shown that chlordiazepoxide produces a deficit in delayed conditional discrimination (Tan *et al.* 1996) of the same form as that produced by Alzheimer's disease in delayed matching-to-sample (Money *et al.* 1992). However, there has been very little study of the possible amnestic effects of buspirone, and data that do exist have not shown any (Lucki *et al.* 1987). Since $5HT_{1A}$ agonists (unlike classical anxiolytics) do not impair delayed matching-

to-sample (Cole *et al.* 1994; compare with Cole and Hillmann 1994), it may be that the amnestic effects of these drugs are more limited even than those of the classical anxiolytics. With regard to the latter compounds, a weakness in current methods of testing them is that they may have involved some element of state dependence (McNaughton 1985); for example Lucki *et al.* (1987) used subjects who were drug-free prior to a single-dose test of the drugs. Firm conclusions, therefore, remain hazardous; but it seems likely that the classical anxiolytics produce a degree of amnestic effect that fits with our general prediction. Animal tests suggest 'that the anxiolytic and amnestic effects of the benzodiazepines are closely related' (Graeff *et al.* 1993, p. 67). However, with long-term, high-dose use, diazepam can have effects on human memory even when its effects on anxiety are no longer evident (Weess *et al.* 1994).

In sum, the data considered in this section suggest that, as expected, anxiolytic treatment results in at least mild hippocampal-like changes in mnemonic function. Consistent with the appearance of theta activity throughout the hippocampal formation and the action of the drugs on theta activity, these effects extend to the more stimulus-oriented tasks which, in the theory, depend more on entorhinal cortex than on hippocampus proper. Thus the effects of anxiolytic drugs appear to be nearly as 'wide' as those of large septo-hippocampal system lesions, but they are not so 'deep'.

These data, then, reinforce our earlier conclusion that 'the septo-hippocampal system, among its functions, includes participation in the behavioural inhibition system; and the behavioural inhibition system, among its neural structures, includes the septo-hippocampal system' (Gray *et al.* 1978). This 'partial overlap' hypothesis leaves the septo-hippocampal system free to subserve such relatively emotion-free activities as learning lists of words in a psychological laboratory; provided, of course, that it has not been pre-empted by matters of greater urgency. In such calmer moments it is likely that the function of the septo-hippocampal system is dominated by its relations with the neocortex, and in human beings perhaps especially by language systems. Under these circumstances the cortically oriented picture of the hippocampus painted by Cohen and Eichenbaum (1993) may be close to the mark.

Our current theory can specify the partial overlap of the septo-hippocampal and behavioural inhibition systems more precisely. We have already noted that the action of anxiolytic drugs on the septo-hippocampal system is 'partial' in a quantitative rather than qualitative sense. That is, we see anxiolytic drugs as affecting the septo-hippocampal system through their disruption of the control of theta activity. This disruption will affect all levels of the septo-hippocampal system, since all receive theta input. Indeed, we have defined the septo-hippocampal system as that set of structures which receive this input. However, the disruption cannot be equated with total hippocampal dysfunction. First, anxiolytic drugs do not eliminate theta activity, they merely alter aspects of its control (Appendix 5). Second, even elimination of theta activity does not eliminate all hippocampal processing (Appendix 9). Anxiolytic drug action, therefore, produces a modest, albeit global, dysfunction of the septo-hippocampal system which, in the theory, we view as a reduction in the *acuity* of hippocampal processing. This suggests that one possible source of pathological anxiety is an increase in such acuity, or in some other process which contributes to excessive functional output from the septo-hippocampal system, producing a dysfunctionally high increase in negative cognitive bias.

Inasmuch, then, as the septo-hippocampal system is involved in cognitive and memorial processing, a consequence of our theory is that pathological anxiety itself is likely to result, at least in some cases, from abnormal cognitive and mnemonic processing (McNaughton 1997). As we consider in more detail shortly, this brings the anxiety aspects of our theory quite close to some recent cognitive theories of generalized anxiety (Eysenck 1992a,b; Mathews 1993). 'Cognitive dysfunction' suggests affective neutrality and a focusing of hippocampal processing on cortical information. But when emergency threatens, the messages received from older structures located in the brain stem take precedence and a fundamentally cognitive dysfunction can have, nonetheless, affective consequences.

11.2 ANXIOLYTIC ACTION AND THE AMYGDALA

There are appropriate receptors to allow direct actions of all classes of anxiolytic within the septo-hippocampal system. However, our theory depends, at present, only on the known direct actions of these drugs on the hypothalamus (Appendix 5), the locus coeruleus, and the raphe nuclei (Appendix 10). All of these direct subcortical actions indirectly produce impairments in hippocampal electrophysiology; and the current evidence suggests that these impairments are sufficient to account for all the hippocampal-like effects of the drugs (Chapters 4 and 6). The hippocampal-like actions of anxiolytics appear to be the result of synergistic multiple direct sites of action, which suggests that the systemic effects of the drugs could include actions also on structures outside both the septo-hippocampal system and the modulatory systems that impact upon this.

As detailed in Chapter 6 (and Appendix 8), there is no evidence that septal or hippocampal lesions eliminate the increment in arousal elicited by conditioned aversive stimuli (Dickinson 1974), although this effect is clearly produced by the anti-anxiety drugs. In the specific case of fear-potentiated startle, there is very strong evidence that the anxiolytics (both novel and classical) produce their effects by a direct action on the amygdala and that dorsal hippocampal lesions are without effect (McNish *et al.* 1997). Thus a detailed specification of 'partial overlap' must include, not only effects of septo-hippocampal damage not fully reproduced by anxiolytic drugs, but also anxiolytic effects not reproduced by lesions of the septo-hippocampal system at all. This consideration leads to a significant addition to the theory of the first edition: anxiety is held to require activity not only in the septo-hippocampal system but often also in the amygdala. It is the amygdala which, in the theory, codes for fear (i.e. cue-specific avoidance behaviour) and implements the 'increase arousal' output of the behavioural inhibition system, one that cannot be attributed to the operation of the septo-hippocampal system. We shall have reason, below, to exclude the periaqueductal grey and hypothalamus from the mediation of anxiety. One tentative conclusion so far, then, is that anxiety could be viewed as a specific state arising from the interaction of the amygdala with the septo-hippocampal system.

On this view, pure fear and pure frustration (the 'purity' arising in each case because of lack of a conflicting approach tendency) would equate with activity in the amygdala (with different patterns of amygdala activity distinguishing these and a number of other emotions); resolution of goal conflicts would equate with functional output from the

septo-hippocampal system; and anxiety (where pure fear or pure frustration is contaminated with conflict) would equate with activity in the amygdala concurrent with output from the septo-hippocampal system. Thus, neither septo-hippocampal nor amygdalar output by themselves can be totally identified with anxiety. This formulation may help to resolve the major discrepancy between the views of LeDoux and Davis, on the one hand, and the theory of the first edition of this book, on the other. It also makes sense of the fact that the amygdala and hippocampus are very tightly linked together. Their connections are not reciprocal, but the inputs from each to the other represent a particularly short recursive loop of a type that is fundamental to our theory of the interactions of the hippocampal formation with all motor-programming areas. There are also further possible sites of coordination between the activities of the hippocampus and the amygdala. It is of particular potential interest that both project to the bed nucleus of the stria terminalis. This structure shares with the amygdala descending connections to lower levels of the defence system and to nuclei that control the whole gamut of autonomic responses that are involved in anxiety; and Davis (1998) has shown that it is involved in the potentiation of the startle response that ensues upon exposure to a diffusely threatening environmental stimulus (bright light). Based upon these observations, Davis suggests that the bed nucleus may play a critical role in anxiety. Attractive as this suggestion is, it requires much further research, as do the interactions between the hippocampal and amygdalar projections to the bed nucleus.

11.3 HUMPTY DUMPTY HAD A GREAT FALL

We must be careful, however, not to fall back into the trap of reifying a 'seat of anxiety' (namely, the hippocampo-amygdalar circuit). It is better to view anxiety as a particular pattern of hippocampal-amygdalar interaction involving also many other areas (see also LeDoux 1996). Emotions such as anxiety can derive their apparent coherence not from the presence in the brain of any single centre coordinating the observed responses, but from the fact that all the observed reactions subserve a common function (or, rather, share a common teleonomy; Pittendrigh 1958, cited by Staddon 1983). There can, then, be many separate 'rules of thumb' (Krebs *et al.* 1983; and see Chapter 5, Section 5.1) which have evolved quite independently and which may have no neural overlap (McNaughton 1989, Section 12.1), but which are correlated because of environmental regularities and so provide the impression of a seamless, higher-order, control system. This appears, in particular, to be the case in relation to the correlated occurrence of, on the one hand, the arousal and autonomic components of activation of the behavioural inhibition system and, on the other, its behavioural inhibition and attention outputs. For, as concluded earlier in the book, the former are coordinated in the amygdala, the latter in the septo-hippocampal system; and, at present, we see no way (other than via environmental regularities of the kind indicated above) in which these two sets of activities are themselves further coordinated.

One of us (McNaughton 1989) has argued strongly for the 'rules of thumb' point of view as applied to emotion in general. Nonetheless, we are forced to that view here with reluctance, since it is undoubtedly less parsimonious than the position adopted at the

time of the first edition. We were then able to identify a relatively unified neural equivalent of the behavioural inhibition system, one that furthermore provided a similarly unified basis for anxiolytic drug action. According to that position, threat-induced increases in the activity of monoaminergic (noradrenergic and serotonergic) afferents to the septo-hippocampal system formed the basis of anxiety (i.e. enhanced processing in the latter system, directed towards the source of threat); and this pattern of activation could be detected by accompanying changes in the electrophysiology (specifically, in the functioning of the hippocampal theta rhythm) of the septo-hippocampal system that were produced in this same monoaminergically mediated manner. Correspondingly, there was a unified basis of anxiolytic drug action, in that these compounds reversed threat-induced increases in monoaminergic activation via action at $GABA_A$ receptors located on both noradrenergic and serotonergic cells. Environmental regularities also formed part of the story, but in a neurologically realized manner. In particular, the participation of GABAergic synapses in inhibitory feedback loops, coupled with the fact that the classical anxiolytics act only at synapses where GABA is already being released, has the consequence that these drugs pick out and increase inhibition in just those circuits that, due to environmental conditions, are currently active: i.e., under conditions of threat, circuits that underlie anxiety. Thus, to the extent that, in 1982, we relied for the unity of anxiety upon a loop through the environment, we could nonetheless see a neurally based rationale for its existence.

Our present position does not abandon any of the specific mechanisms outlined above. It does, however, add to them in ways that weaken the 1982 model.

First, as noted above, we now need to pay due attention to the amygdala. The role of this structure in coordinating the arousal and autonomic outputs of the behavioural inhibition system is unaffected by hippocampal lesions, and therefore cannot be secondary to events that first take place in the septo-hippocampal system. It follows that the mechanisms described in the previous paragraph can no longer be seen as fundamental to all aspects of anxiety, only to some of them. Our Humpty Dumpty, then, has had a great fall. A possible way to put him together again might emerge, if the role of the amygdala in coordinating the arousal and autonomic outputs of the behavioural inhibition system were secondary to threat-induced increases in the monoaminergic inputs it receives; but we know of no evidence at present that supports this hypothesis.

In any case, even if this possibility were backed by experimental evidence, there are two further ways in which we are forced to abandon the unity of our 1982 position, ways that are unrelated to the role played by the amygdala in anxiety. As noted in Chapter 9, the actions of the anxiolytics upon the hippocampal theta rhythm are twofold: alterations in the sensitivity to theta frequency of the septal input to the hippocampus, and reductions in the frequency of theta resulting from a specified intensity of afferent input to the septal area. We now know that only the first of these electrophysiological signatures of anxiolytic drug action reflects dampening of ascending monoaminergic activity; the second, mediated by drug action in the supramammillary nucleus, appears to be a completely independent effect.

The third set of findings that weaken our 1982 position leads to similar inferences. We have made much, and justly so, of the fact that the novel anxiolytics have the same behavioural and electrophysiological effects as the classical anxiolytics. However, while

this pattern of findings strengthens the case for regarding anxiolytic drug action as an excellent guide to the neurology of anxiety, it simultaneously eliminates enhanced transmission at $GABA_A$ receptors (which the novel anxiolytics do not produce) as a necessary step in such action. But, if so, we can no longer look to the special role played in the nervous system by this receptor, and by the allosterically coupled receptors upon which the classical anxiolytics act, for a neurally realistic explanation of the role played by the environment in keeping together the congeries of brain structures that must now be included under the rubric of 'the anxiety system'.

We can no longer affirm, therefore, that the actions of the anxiolytic drugs which led us to the parsimony of our model of the behavioural inhibition system, at the level of the conceptual nervous system, reflect a single mode of action of these compounds in the real nervous system; and, correspondingly, given the central role that anxiolytic drug action has played in our overall scientific strategy, we can no longer affirm that, in the real nervous system, there exists a neurally unified set of structures that match neatly onto the behavioural inhibition system. That, clearly, represents a major loss in the structural simplicity of our model. Future research may identify alternative ways in which to put Humpty Dumpty together again (we return to this issue in the final two chapters). For the moment, however, even as a congeries united only by 'rules of thumb', we believe that our approach succeeds in identifying and analysing the set of structures by which the brain handles anxiety better than any of its (generally much simpler) rivals.

11.4 ANXIOLYTIC ACTION AND THE NEOCORTEX

We have related a certain type of interaction of the septo-hippocampal system and amygdala to anxiety, but is this perhaps just one of the rules of thumb contributing to this state? This is indeed likely to be the case. We have distinguished 'anxiety' from 'fear' as being the case where the animal must enter into, as opposed to exit from, a dangerous situation (Chapters 2 and 3). This involves risk analysis, anticipation, and many other complex processes. Furthermore, anxiety can derive as easily from the anticipation of loss of reward as from the anticipation of punishment. Again, this suggests complex neural organization. Finally, a primary feature of anxiety is the role played by the anticipation of both appetitive and aversive events and the need to trade the one off against the other. This suggests that some cases of anxiety could involve the highest levels of the nervous system.

Recall one of the key architectural features of the various systems considered at different points in this book. With the exception of the aminergic systems (which originate in small nuclei and are unlikely to process complex information), these are all hierarchically organized with 'quick and dirty' components at the bottom and 'slow and sophisticated' components at the top (but still integrated with, and often operating through, the lower levels). In the case of anxiety, we can see the septo-hippocampo-amygdala loop as the lowest level of the system (with the hypothalamus and periaqueductal grey representing separate fear- or panic-oriented systems rather than anxiety proper). Our theory, then, is compatible with the possibility that higher areas, such as prefrontal and cingulate cortex, can also be involved in anxiety. However, the implied higher aspects

of anxiety controlled by these areas appear to be immune to anxiolytic drugs (hence the occasional clinical need for psychosurgery, as considered above).

Provided that we can equate at least some forms of anxiety with the processes affected by anxiolytic drugs, we have now achieved a preliminary neurology of anxiety, the finer details of which are determined by our theory of the septo-hippocampal system (Chapter 10). This neurology of anxiety derives largely from animal experiments, psychological, physiological, and pharmacological. But our aim has always been to provide a new perspective on the manifestations of human anxiety, and particularly pathological anxiety. It is time to see how well we have achieved this aim. As just noted, there is some anxiety, with very similar psychological properties to anxiolytic-sensitive anxiety, which is not sensitive to these drugs but responds to frontal and cingulate therapeutic surgery. There is also a range of entities (including phobia, panic, and obsessions) which figure in other typologies of anxiety, but which we exclude from our own definition on both behavioural and pharmacological grounds. We treat these below as 'anxiety-related' rather than 'anxiety' disorders in an attempt to keep clear the distinctions we wish to make, while also retaining the general grouping of disorders found in the conventional classifications. We need to provide, therefore, a basic typology of the 'anxiety disorders' which maps the clinical phenomena to the elements of our theory. To achieve this, we must translate our theory of the septo-hippocampal system, and its implications for the behavioural inhibition system, into a full psychological theory of anxiety and anxiety-related disorders.

11.5 THE CLINICAL PSYCHOLOGY OF ANXIETY

Here we provide only a brief descriptive overview of the phenomena of clinical anxiety. Various analyses of normal (non-clinical) anxiety are provided in Chapters 2 and 3. We shall argue later that, at the level of symptoms, there is no fundamental difference between normal and pathological anxiety. One reason for starting with anxiety *disorders*, however, is the belief that they are extreme manifestations of aspects of normal anxiety. If this is true, their exaggeration may provide particularly clear clues to the nature of the anxiety which, in everyday life, is likely to be intermixed with many other emotional states. Another reason is to emphasize the practical importance of what might otherwise seem a rather academic exercise. 'It is estimated that nearly 15% of the general population in the US has suffered from at least one DSM-III anxiety disorder at some point in their lives. . . . The burden of anxiety disorders extends beyond the direct costs of treatment to the indirect costs of impaired social functioning. People with anxiety disorders suffer from considerable social morbidity—they have elevated rates of financial dependence and unemployment, substance abuse and dependence' (Leon *et al.* 1995, pp. 19, 21–22). Thus, we present below detailed discussion and analysis of specific clinical phenomena, considered in the context of the theory.

Behavioural symptoms of any one specific central state or kind of disorder do not come with convenient labels distinguishing them from those of other states or pathological conditions. Indeed, the problem of the classification of behavioural disorders is one of the most difficult that psychiatry faces (see for example Marks 1988). Many different approaches to this problem have been proposed, but will not be discussed here.

We shall simply restate the fundamental axiom of our approach: that we must look at data first and attempt definitions only later. This approach has resulted in two main biases. Where proneness to anxiety is concerned, we favour the dimensional approach advocated, for example, by Eysenck (1960; see Chapter 12). But, where syndromes are concerned, we favour a neurological approach, that is the attempt to identify the particular brain systems concerned. These brain systems are related to syndromes in both of two distinct but mutually reinforcing ways: (1) those in which dysfunction gives rise to the disorder; and (2) those in which activity underlies specific symptoms. It is in the spirit of this latter, neurological approach that even clinicians who hold that 'the inheritance of trait anxiety accounts for the familiality of anxiety disorders' find that treatment 'programs are still diagnosis-specific. Accurate diagnosis and assessment of each patient is critical for effective treatment' (Andrews *et al.* 1994, pp. 8, 14). One challenge for our theory, then, is to reconcile these 'dimensional' and 'neurological' approaches, despite their constituting, apparently, polar opposites to each other; we return to this issue in Chapter 12.

Detailed descriptions of the syndromes and symptoms of anxiety can be found elsewhere (Marks 1969, 1988; Mayer-Gross *et al.* 1969; Beech 1974; Lader 1975; Rachman and Hodgson 1979; Noyes *et al.* 1988). We take as our starting point, DSM-IIIR, probably the most widely used recent method of diagnosing patients as suffering from anxiety disorder and of dividing them into subtypes. (As will be seen, starting with DSM-IIIR, does not mean that we agree with it, nor with its recent successor DSM-IV, nor even with their basic philosophical approaches. But we have to start somewhere.) The DSM-IIIR definition of anxiety is as follows:

> apprehension, tension, or uneasiness that stems from the anticipation of danger, which may be internal or external. Some definitions of anxiety distinguish it from fear by limiting it to anticipation of a danger whose source is largely unknown, whereas fear is the response to a consciously recognized and usually external threat or danger. The manifestations of anxiety and fear are the same and include motor tension, autonomic hyperactivity, apprehensive expectation, and vigilance and scanning.
>
> Anxiety may be focused on an object, situation, or activity, which is avoided (phobia), or may be unfocused (free-floating anxiety). It may be experienced in discrete periods of sudden onset and be accompanied by physical symptoms (panic attacks). When anxiety is focused on physical signs or symptoms and causes preoccupation with the fear or belief of having a disease, it is termed hypochondriasis. (DSM-IIIR, American Psychiatric Association 1987, p 392.)

It is interesting to note that no part of this definition necessarily implies pathology. Pathology lies in the inappropriateness of the timing or the intensity of anxiety, given the eliciting stimuli.

At the syndrome level, DSM-IIIR distinguishes between a wide variety of subtypes of anxiety disorder, in addition to differentiating anxiety disorders proper from those organic, mood, and psychotic disorders which can be accompanied by symptoms of anxiety. It itemises: separation anxiety disorder; simple phobia; panic disorder and/or agoraphobia; social phobia [avoidant disorder of childhood or adolescence]; obsessive compulsive disorder; generalized anxiety disorder [overanxious disorder]; post-traumatic stress disorder; adjustment disorder with anxious mood; and 'anxiety disorder not otherwise

specified'. (The diagnoses in square brackets are made on the basis of an upper age limit.) A question considered below is how far these distinctions represent variations resulting from disorder in a single underlying system and how far they reflect disorders in distinct neural systems.

At the level of symptoms it is also common to distinguish between behavioural, cognitive, and physiological aspects of anxiety (Lang 1970; Rachman 1978). This division is useful, so long as it is recognized that it cuts across the division of systems in the brain implied by the DSM-IIIR categorization of syndromes.

11.6 BEHAVIOURAL ASPECTS OF ANXIETY SYNDROMES

Behaviourally, the chief symptom of the phobias is avoidance of the phobic stimulus. If this avoidance is to be classified as pathological, the phobic stimulus should be one which is not objectively dangerous or the avoidance behaviour should be incapacitating. In some cases the phobic stimulus may be idiosyncratic (e.g. cats, a special pattern of wallpaper) and is likely to depend heavily on a specific conditioning history. In many cases, however, the phobic stimulus will evoke fear or anxiety (but not at a level which produces behavioural incapacity) in the normal population (e.g. spiders, snakes, the dark). Here, it can be argued that there is some degree of genetic preparedness (but see Davey 1995). In most cases of simple phobia, the disorder appears to lie in the unusual magnitude of the fear reaction rather than in any peculiarity of its intrinsic nature or eliciting stimulus. We argue later that phobia proper should be categorically distinguished from anxiety.

Agoraphobia (avoidance of public places) and social phobia are particularly crippling and might seem at first glance to be only extreme and generalized forms of simple phobia. However, there is now good reason to view social phobia as a quite distinct subtype of anxiety (Mattick 1990). Similarly, while there are specific place phobias which appear to be simple phobias with a specific place as the phobic object, they should not be confused with generalized agoraphobia. This involves 'a cluster of public places as the chief cues for anxious avoidance', something which is emphasized more in the ICD-10 classification scheme than DSM-IIIR (Marks 1987, p. 12). We argue later that agoraphobia and social phobia would be better named 'agoranxiety' and 'social anxiety' respectively.

Agoraphobia commonly occurs together with panic attacks. Indeed, on one view, agoraphobia is always a means of avoiding, or avoiding the consequences of, panic attacks (e.g. Friend and Andrews 1990; Joyce and Oakley-Browne 1990). In contrast, Marks (e.g. 1987) argues that panic is the result of extreme phobia produced by any of a variety of stimuli, and is more prominent in agoraphobia only because the fear stimuli are more difficult to avoid. In support of Marks' view, agoraphobia proper seems to precede panic in some cases (Lelliott *et al.* 1989). Moreover, as many as 1 in 20 of a mixed panic and/or agoraphobia sample can be diagnosed as agoraphobic without history of panic disorder, even after exclusion of phobias of specific stimuli such as crowds, tunnels, etc. (Goisman *et al.* 1994). We accept, then, with Marks, that high levels of anxiety can precipitate panic (as can a variety of other stimuli, such as a high partial pressure of carbon dioxide), and that agoraphobia can precipitate panic attacks in panic-prone individuals. Agoraphobia, furthermore, is not the only such condition: generalized

anxiety disorder, too, can on occasion predispose to panic disorder (e.g. Brawman-Mintzer 1994). However, we accept also that, in many cases, panic *disorder* (i.e. the *inappropriate* occurrence of panic attacks) gives rise, through conditioning in individuals with a predisposing personality (i.e. neurotic introverts, an issue to which we return in Chapter 12), to agoraphobia (see review by Gorman *et al.* 1989).

Taken together, these two views imply that agoraphobia as a clinical entity can, on occasion, involve positive feedback between anxiety and panic in a vicious circle (independent of whether it is panic or anxiety that is causally primary). Despite this potential, and indeed common, positive feedback, we distinguish (see also Okasha *et al.* 1994) panic itself (and hence panic disorder) from anxiety (and anxiety disorder). Certainly, the two are not necessarily linked: there are cases of panic attacks that occur only when the patient is asleep, but not when anxiety is experienced in the waking state (Rosenfeld and Furman 1994); and also indications that panic disorder is associated with autonomic dysfunction (e.g. Yeragani *et al.* 1994), whereas no evidence exists of such an association with anxiety. Although our detailed neurology is slightly different from that of Gorman *et al.* (1989), the picture we ultimately paint is consistent with their view that: 'there are three distinct components of the illness panic disorder—the acute panic attack, anticipatory anxiety, and phobic avoidance. These three clinical phenomena arise from excitation in three distinct neuroanatomical locations, respectively: the brainstem, limbic lobe and prefrontal cortex. Reciprocal innervation among nuclei in these three centers explains the genesis of the disease and its clinical fluctuations over time' (Gorman *et al.* 1989, p. 149).

Obsessive–compulsive disorder appears to parallel panic disorder in its general aetiological relationship to anxiety. Obsessions (described more fully in the next section) and compulsions appear initially as occasional intrusions into the mental and behavioural life of an individual who perceives them as irrational. The disease is progressive, with an increasing frequency of occurrence of symptoms, and these ultimately become socially incapacitating. In the majority of cases there is a ritual behavioural act ('compulsion') which seems reasonable except in its frequency. The washing of hands to avoid infection is an example: it is the *repetition* for hours on end (to the point sometimes of producing severe dermatitis) that appears unreasonable, not any single act of washing. It should be emphasized that the compulsive behaviour does not itself constitute anxiety. Indeed, it has been argued (e.g. Rapoport 1989) that both obsessions and compulsions are the result of a disorder of higher-order motor control programs, not unlike the disordered simple movements which are symptoms of neurological syndromes such as Huntington's chorea. *Prevention* of the compulsive ritual, however, gives rise to a sharp increase in the subjective experience and physiological signs of anxiety (Roper and Rachman 1976). Despite the increase in anxiety, response prevention can be used as a behavioural means of controlling compulsive behaviour. But this technique seems much less successful in controlling the associated obsessions. It is a moot point whether the obsessions and compulsions are the result of anxiety, which they effectively decrease, or whether the anxiety is simply the consequence of failure to achieve safety. An everyday ritual such as locking your front door is likely to be performed in a habitual fashion, with no cognitive or bodily signs of anxiety. It is only when you realize that you have forgotten to lock the door (or you are prevented from locking it) that you feel anxious.

So, obsessive–compulsive disorder (like panic disorder) may not be primarily an anxiety disorder, but rather generates symptoms which, as they become more difficult to accommodate, can give rise to anxiety. This would be particularly likely in individuals with a neurotic introvert personality. Equally, it is clear that increases in anxiety can give rise to at least an increase in the frequency of obsessions and compulsions (paralleling the increase it can produce in the frequency of panic attacks). The repetitive checking of a toddler by a parent, for example, can seem obsessive to a non-parent; in this type of case, we attribute the increased checking to anxiety, rather than the other way round. As noted for panic and agoraphobia, there is then the possibility that, in pathological cases, interactions of this kind could give rise to a vicious circle; and also that there may be different cases that are primarily anxious or primarily obsessional, respectively, both giving rise, however, to similar patterns of mixed anxiety and obsession.

11.7 COGNITIVE ASPECTS OF ANXIETY SYNDROMES

The cognitive aspects of simple phobias are not dramatic, and in practice often boil down to the self-report of fear or anxiety, or merely a tick placed at an appropriate point on an adjective checklist (although such self-reports are seldom obtained at the height of an intense attack).

By contrast, obsessive–compulsive disorder has a rich cognitive phenomenology. Obsessions consist of 'intrusive, repetitive thoughts, images or impulses that are unacceptable and/or unwanted and give rise to subjective resistance; the person finds them difficult to dismiss or control' (Rachman 1978). Given the relative ease with which compulsions can be suppressed and the relatively refractory nature of obsessions, it is possible that the latter constitute the primary aspect of the disorder. Rapoport's (1989) view that the obsessions are a high-level motor disorder is tantamount to the idea that a belief (e.g. that one is dirty) is a high level of motor command. The analogy is with choreic movements. These are spontaneously occurring, isolated elements of normal movements which occur as the result of neurological abnormality in the basal ganglia. In a similar manner, Rapoport views obsessions as the spontaneous occurrence of isolated elements of normal planning, including the accompanying cognitions. On this view, the cognitions most often reported in the disorder may be precursors to anxiety rather than reflecting anxiety itself.

Panic according to Marks (1988b, p. 115) 'denotes a surge of intense anxiety or similar discomfort'. However, it may be associated with cognitions such as 'fear . . . of dying, losing control, or going mad' (Marks 1988b, p. 120).

Generalized anxiety disorder, in contrast to all of the above, can be viewed as anxiety, the level of which is out of proportion to the level of any perceived sources of threat. It is, in a sense, anxiety looking for a threat to perceive rather than anxiety resulting from a perceived threat; and 'during the course of the disorder, the focus of worry may shift from one concern to another' (DSM-IV; American Psychiatric Association 1994, p. 433). It has recently been suggested that the critical pathology in this disorder lies in the functioning of cognitive, particularly working memory, systems (Eysenck 1992; Eysenck and Calvo 1992) or the control of attentional resources (Mathews and Macleod

1994). While the cognitive processes we invoke are different, our theory has much in common with these views, sharing in particular the idea that generalized anxiety is primarily a cognitive disorder.

In all the anxiety-related disorders there is a tendency to attend particularly strongly to environmental stimuli or events that the patient finds particularly threatening. Thus, in panic disorder, the patient is particularly aware of bodily change, e.g. increased heart rate or feelings of breathlessness, which he or she interprets as signs of impending illness or death; or, in social phobia, patients tend, when anticipating a social interaction, selectively to 'retrieve negative information about how they may appear to others' (Clark 1997), especially in the form of negative images of themselves as seen from an outside observer's perspective (Hackmann *et al.* 1998). This pattern of enhanced attention to threatening stimuli is, of course, exactly what would be predicted as a consequence of excessive activity in the behavioural inhibition system, since one of the major outputs of this system has just this function. At the time the first edition of this book was published, however, there was little if any evidence that heightened attention to threatening stimuli plays the major role in anxiety that has since been established on a strong empirical basis (Mathews and Macleod 1994). We take this theme further in Section 11.10 below (and see Chapter 13 for cognitive approaches to therapy).

11.8 PHYSIOLOGICAL ASPECTS OF ANXIETY SYNDROMES

In all the anxiety syndromes there are easily detectable peripheral physiological signs: raised skin conductance with an increased rate of spontaneous fluctuations in conductance level, raised heart rate, increased blood flow, etc. (Lader 1975). When panic attacks are present, there is a chronically raised level of activity in the autonomic nervous system between as well as during attacks (Lader and Wing 1966; Lader 1967). Some chronically raised autonomic activity is also seen in social but not simple phobia (Lader 1967), although in this condition increased responsiveness of the autonomic system to simple physiological challenge may be more important than the increased level of anxiety *per se* (Sten *et al.* 1994a).

11.9 THE THEORY OF ANXIETY

We began the task of constructing a theory of anxiety armed, at the psychological level, with an ethological dissection of defence (Chapter 2) which distinguished immediate from potential threat, the latter mapping onto the learning-theory concept of behavioural inhibition (Chapter 3). The ethological analysis will prove important for the way in which we deal with comorbidity, and also for its functional equation of innate and acquired anxiety stimuli. But it is the learning-theory concept of behavioural inhibition which is fundamental to our view of anxiety as such. This concept was based largely on the results of purely behavioural experiments (Gray 1975), although the behavioural effects of anxiolytic drugs also played an important adjunctive role in its early development (Gray 1967, 1977). Most of this book has been concerned with a search for, and

a detailed analysis of, the neural structures which might subserve the various functions allotted to the behavioural inhibition system. It may be surprising that, as a result of this search, it has become necessary to define more sharply and deeply the *psychological* content of the concept of the behavioural inhibition system; but this fruitful interaction is fundamental to the power of the super-discipline of behavioural neuroscience, where each discipline (psychology, physiology, pharmacology, neurology, anatomy, molecular biology, etc.) can force changes in the others.

Both the ethological analysis of Chapter 2 and the psychological developments enforced by the neural details of our theory have had the consequence that the original model of the behavioural inhibition system (Gray 1975) has undergone considerable evolution, although the major postulates that it encapsulated need not be abandoned. This system can still be regarded as responding to stimuli which warn of punishment, stimuli which warn of non-reward, novel stimuli, and innate anxiety stimuli. But, conditioned or unconditioned, these can now all be seen as special cases of the more general functional class of stimuli which warn of potential negative affective events. Equally importantly, it is a key feature of these stimuli that the organism has some reason to approach them. This characterization makes explicit something that was only implicit in the first edition of this book. It is the *evaluation* of such stimuli that is critical to their effect, in other animals as well as human beings. In turn this implies that, particularly in human beings, we must take account of high-level cognitive mechanisms, although these are probably not fundamentally different from those at work in other animals (see Chapter 8, Section 8.1). In the latter context, we need to take into account recent developments in conditioning theory that go well beyond older S–R ideas (see for example Mineka and Zinbarg 1996). We can also still characterize the outputs of the behavioural inhibition system as comprising inhibition of ongoing behaviour, increased attention to environmental stimuli, and increased level of arousal. But the need to specify how the brain might discharge these functions has forced us to define more clearly the kinds of information processing it must undertake. Thus, to see how adequate an account of anxiety is now provided by the notion of 'activity in the behavioural inhibition system', we must first recapitulate certain features of the theory of the septo-hippocampal system developed in Chapter 10.

As argued there, the chief function of the latter system is to compare, quite generally, currently primed goals with each other and with expectations. An individual goal is defined in terms both of the stimuli (available and expected) to which a response can be addressed and of the class of motor programs (or 'plans'; Miller *et al.* 1960) which can achieve the goal. So long as there is no goal conflict, the system does nothing directly to influence ongoing behaviour. But that does not mean that, at such times, it is inactive. On the contrary, it is kept extremely busy, receiving information from a variety of areas primarily concerned with the processing of stimuli, or with the programming of responses, but which all have in common the capacity to direct the upcoming behaviour of the animal. In this comparator or monitoring capacity ('just checking') the behavioural inhibition system is continuously active (as shown by single-cell or slow-wave recording in the hippocampus); but it produces functionally significant output, and hence controls behaviour, only under special conditions. These conditions ('conflict') are those where multiple goals are primed, either because an unpredicted environmental

event occurs, or because a predicted event fails to occur, or because there are equal incompatible tendencies (approach–avoidance is the most important for the current chapter, but approach–approach or avoidance–avoidance are also possible).

Conflict between goals causes the system to enter 'control mode'. Under these conditions, the system at once interrupts ongoing motor behaviour in the manner that initially gave rise to the concept of the 'behavioural inhibition system'. Of particular importance for our understanding of anxiety, not only is behaviour addressed to the current prepotent goal inhibited, but the stimuli associated with that goal are tagged 'faulty' and called in for careful inspection. The 'faulty' tag has the theoretically crucial effect of increasing the negative affective bias of stimulus assessment mechanisms (Fig. 1.7), allowing resolution of the conflict by the resultant inhibition of those goals with the greatest negative associations. Thus, when the stimuli next occur (and activate the representation of the goal), approach behaviour is interrupted so that the system can determine whether alternative programs lead to more satisfactory outcomes. 'Determine' here is used in a distributed sense. In fact, conflict causes the septo-hippocampal system to send very simple feedback signals to the various systems, each of which could potentially take control of behaviour. These feedback signals increase the valence of affectively negative associations and will reduce the level of activity (priming) in those areas with the greatest negative associations. In many cases this will be sufficient to leave one area with significantly greater priming than the others, and this will then take control. If recursive interaction between the different goal representations and the septo-hippocampal system does not fairly swiftly produce a clear winner, the system itself generates exploratory and investigative behaviour (risk analysis), especially checking for sources of environmental threat. As noted in Section 11.2 above, an important elaboration of the theory in the present edition is the view that, while the business of the septo-hippocampal system is the resolution of conflict between goals in the most general sense, the more specific case of defensive conflict is presumed to involve concurrent activity in both the septo-hippocampal system and amygdala.

A valuable perspective on the clinical situation is provided by the pharmacotherapy of anxiety and anxiety-related disorders. As we have seen, classical and novel anxiolytics share neither direct receptor actions nor side-effects. Their only common clinical action is on anxiety. Comparison of the effects of these chemically quite different types of drug gives us, then, a scalpel with which to dissect the components of anxiety disorders. Even with respect to anxiety as defined by DSM-IIIR, this dissection simplifies the picture which would have been obtained 10 years ago. Panic disorder proper can no longer be equated with the anxiety which often accompanies it, exacerbates it, or produces it, since panic is not affected by buspirone; atypical depression (with its anxiety-like somatic symptoms) cannot be equated with anxiety, since it is insensitive to the benzodiazepines; the same is probably true of most conventional depression (with antidepressant benzodiazepines like alprazolam being exceptions which prove the rule); and, like panic, obsessive–compulsive disorder can be viewed as a condition which both gives rise to and is exacerbated by anxiety, but which is nonetheless pharmacologically distinct, as shown by drugs such as clomipramine that are particularly effective at treating it. Finally, because they are insensitive to anxiolytic drugs, we can exclude specific phobias from our analysis, and identify them with pure fear rather than anxiety.

At a more general, psychological level, our analysis leads to the conclusion that, in the core state of anxiety from which panic, phobias, obsessions, and compulsions have been stripped out, the central dysfunction consists in overactivity in the behavioural inhibition system, i.e. in the interlinked set of neural structures that mediate the functions of this system, as outlined above.

11.10 CLINICAL ANXIETY AND MEMORY

Resolution of conflict, then, is at the core of the theory of the septo-hippocampal system; and defensive conflict is at the core of the application of the theory to anxiety. But we need also to look at the connection between anxiety (where we started the book) and memory (a key destination to which the hippocampal component of the theory took us).

A final and critical fact, evident only in the last few years, which modifies and simplifies the theory of the first edition is this: the anxiolytic drugs do not change the symptoms of pathological anxiety immediately. This appears to be true even of benzodiazepines, which mask the lack of an initial truly anxiolytic effect with muscle relaxant and euphoriant effects (Appendix 1). However, the common effects of all classes of anxiolytic drugs on the septo-hippocampal system are immediate and show no change with repeated administration (Appendix 5). We conclude from this that the drugs block a mechanism which *maintains and creates* anxious symptomatology, and that the clinical delay results from the time taken to extinguish previously acquired patterns of behaviour and cognition. We must make explicit, then, what was previously only implicit in the 'faulty, needs checking' tag of the first edition: the action of the anxiolytic drugs appears to be on something akin to a memory mechanism. This inference brings us very close to current views of both the functions of the hippocampus and the cognitive aspects of generalized anxiety disorder.

As argued in the previous section, anxiety results from overactivity in, or at least excessive functional output from, the septo-hippocampal system. This should lead to both increased risk assessment behaviour and an excessive negative bias in the assessment of goals, via excessive weighting of the negative associations of stimuli, often resulting in excessive tagging of motor programs as faulty. Of particular interest in this context are recent developments in the psychology of generalized anxiety disorder, which link it with cognitive bias (see Mineka and Sutton 1992). M. W. Eysenck (e.g. 1992a) suggests that this disorder results from a specific fault in working memory; and Mathews (1993; Mathews and MacLeod 1994) emphasizes alterations in the attentional control of cognitive and memorial resources towards the processing of threatening stimuli in anxiety. Our rejection of the idea that the hippocampus is directly involved in working memory and our emphasis, earlier, on interference bring us closer to Mathews' views (see Mathews and MacLeod 1994, pp. 41–2) than to Eysenck's, but it will become clear that our position overlaps substantially with both.

In particular, Mathews and MacLeod (1994) suggest that

> early automatic (pre-attentive) analysis of emotional information is relatively global, and perhaps confined to classifying a stimulus as potentially threatening, but that subsequent processing becomes

increasingly selective, favoring information that matches current concerns . . . Anxious mood (or stressful events) leads high, but not low, trait anxious subjects to selectively encode threatening information. . . . These observations are consistent with the possibility that individual differences in selective encoding represent the cognitive substrate for vulnerability to emotional disorder.

Of particular interest to our theory, such selective encoding does not appear to occur under all circumstances. Rather, there is evidence for

a selective mechanism which is used when at least one threatening and one neutral stimulus are present concurrently . . . Those high in trait anxiety will allocate processing resources to the threatening stimulus rather than the neutral one . . . The selective mechanism cannot be used in situations . . . in which only one stimulus is presented at a time. . . . [Similarly,] when ambiguous stimuli are presented which can be interpreted in either a threatening or a neutral fashion, those high in trait anxiety typically produce a greater number of threatening interpretations than those low in trait anxiety. (Eysenck 1992b, p. 171.)

We appear, here, to be dealing with a predisposing factor (a theme we resume in Chapter 12), which is consistent with excessive output from the behavioural inhibition system and which could over time result in an excessive association of threat with environmental stimuli. Thus,

patients who have recovered from generalized anxiety disorder and normals high in trait anxiety both exhibit greater susceptibility to distraction with threat-related distractors than with non-threatening or neutral distractors. If high trait anxiety can be regarded as predisposing to generalized anxiety disorder, then these findings provide reasonable evidence for the notion that sensitivity to threat distraction forms part of a cognitive vulnerability factor for generalized anxiety disorder. The notion that sensitivity to threat distraction depends primarily on long-lasting structural effects within the cognitive system rather than being a function of the current level of experienced or state anxiety is strengthened by the additional finding that the effects of threat distraction were unaffected by the S's level of state anxiety. (Eysenck and Byrne 1992.)

To summarize, contemporary cognitive theories of generalized anxiety disorder accept that there must be conflict before the relevant cognitive biases are displayed; and they include in those cognitive biases both increased attention to threatening stimuli and increased magnitude in the level of threat perceived in any particular stimulus. In particular where a stimulus is ambiguous, having both threatening and non-threatening aspects, the valence of the former aspects is increased. These propositions are also ones that we accept.

11.11 A FUNCTIONAL TYPOLOGY FOR DEFENCE

In this section we make explicit the typology of defence, and hence anxiety disorders, which has been implicit since our adoption of the Blanchards' analysis of defensive distance, Graeff and Deakin's views on the hierarchical organization of defence systems, and LeDoux's views on the hierarchical organization of the neural systems which detect the adequate stimuli for defence (Chapters 2 and 6). We first emphasize the categorical

division of threat stimuli which can be made on purely functional grounds. This is shown in Fig. 11.2 (derived from Fig. 2.3). Threat in general can be divided into two basic types, each with different, indeed in some cases opposed, functional requirements: actual present threat which must be avoided (e.g. the cat in the Blanchards' experiments) and potential threat which must be approached (e.g. the smell of the cat). Each of these, in turn, can be divided into two subtypes, with different (and again opposed) functional requirements: avoidable threat and unavoidable threat.

Each of the four resultant categories can be related to different types of required and, in the Blanchards' analysis, actual behaviour. Actual unavoidable threat, for example, can give rise to undirected escape behaviour or defensive aggression. Actual avoidable threat can give rise to directed escape behaviour. Potential avoidable threat can give rise to active risk assessment. Potential unavoidable threat can lead to a suppression of active coping strategies and a conservation of resources. These different functional classes and, to a lesser extent, the behaviour patterns elicited by them, can be mapped as shown in Fig. 11.2, to the emotions of panic, fear, anxiety, and depression (although we do not deal with the latter condition in this book). (For a detailed discussion of the relation between such functional classes and the output of skeletal, autonomic, and hormonal systems, see McNaughton 1989a.) There are two additional categories in Fig. 11.2 which can be viewed as special cases of unavoidable actual threat and avoidable potential threat, respectively. First, a predator may be definitely present and unavoidable in the sense of flight, but may be perceived as avoidable if attacked. This results in activation of fight behaviour. Second, a potential source of danger (e.g. infection) may be avoidable in the sense that some action can be taken (e.g. hand washing), but there may be no explicit

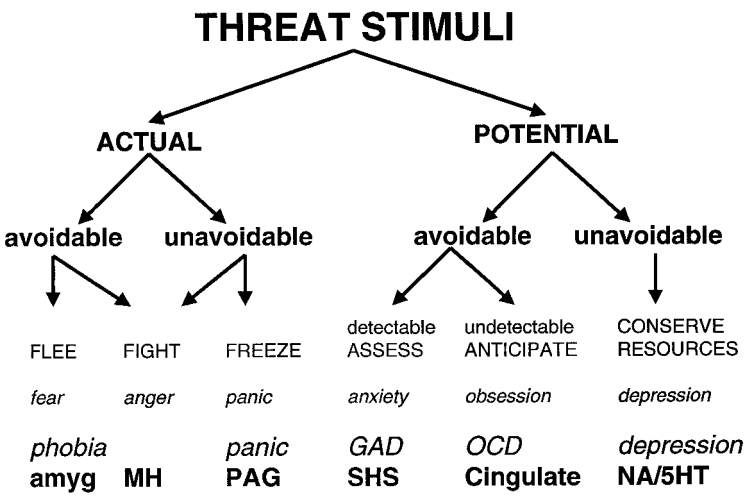


Fig. 11.2 Nature of stimuli (top three rows) and their relation to function (fourth row), emotion (small italics), psychological disorder (large italics), and principal neural system involved (bottom row). GAD, generalized anxiety disorder; OCD, obsessive–compulsive disorder; amyg, amygdala; MH, medial hypothalamus; PAG, periaqueductal grey; SHS, septo-hippocampal system; NA, noradrenaline; 5HT, 5-hydroxytryptamine. For further details see text.

threat stimuli which indicate that the potential danger is present; or, once action has been taken, no explicit safety signals which indicate that it has in fact been avoided. This results in activation of rituals which attempt to anticipate the potential danger. We see the adaptive emotions of panic, fear, and anxiety as also mapping to the pathological conditions of panic disorder, phobia, and generalized anxiety disorder, respectively. Finally, we attribute the *symptoms* of panic to activity in the periaqueductal grey, of phobia to activity in the hypothalamus and amygdala, of anxiety to activity in the septo-hippocampal system and amygdala, of obsession to activity in the cingulate cortex, and of compulsion to activity in the basal ganglia. We also attribute what we term below the *primary* disorders of panic, phobia, etc. to pathological (in the psychological and functional, but not necessarily neurological, sense) hyperactivity in these same brain areas (but note that these distinctions do not map neatly onto current clinical classifications).

Suppose, for the moment, that we have to deal with only the specific cases of panic, fear, obsession, and anxiety. There are many who will be happy with the idea of distinguishing between these cases insofar as they refer to symptoms. However, the suggestion that it should be possible to distinguish between them categorically at the syndrome level may seem, especially to clinicians, a gross oversimplification. One further theory-driven step needs to be taken before we can bring this typology into contact with clinical reality. This step will show that the apparent problem arises from the fact that we must distinguish very carefully between symptoms and syndromes. But how can this be, if we identify, as we did in the previous paragraph, *symptoms* with activity in specific brain structures, and *syndromes* with hyperactivity in those same brain structures?

The key, in our opinion, to a deeper understanding of anxiety and anxiety-related disorders lies in recognising that the various neural structures which control defensive behaviour are strongly and recursively interconnected both neurally and through environmental consequences. This property has already been shown to be important in terms of inhibitory interactions, as in the case of Graeff's explanation of relaxation-induced panic as arising from opposite interactions of 5HT systems with amygdala-controlled fear and periaqueductal grey-controlled panic (Chapter 6). However, it is even more important when we consider the various possible positive feedback interactions. In addition to the specific processing which goes with the need to evaluate and react to specific threatening stimuli, activation of any part of the defence system increases overall arousal and activity in the aminergic attentional systems, and hence the sensitivity of all of the elements of the entire distributed network with which we are dealing. This provides a route, for example, for fear or anxiety to produce a level of arousal sufficient to induce a panic attack as a result of the effective decrease in defensive distance (as defined by the Blanchards, see Chapter 2).

The particular cases we consider in most detail below are those in which the prevention of compulsive rituals engenders an anxiety which was not previously evident, and in which spontaneous panic attacks can elicit anxiety through conditioning. Neural recursive interactions can be presumed to operate similarly in both normal and pathological anxiety. The environmentally mediated ones are likely, however, to be most obtrusive in pathological anxiety. In any case, in most real-life situations, the available stimuli will not in fact be selective for just one of the distinct functional categories we have been considering. The rabbit which sees a fox coming towards it does not necessarily know

whether the fox has seen it. Thus, while the presence of the fox is definite and actual, the threat presented by the fox must be treated as having simultaneously the properties of both actual and potential threat. Thus both fear and anxiety goal-processing systems will be primed for intense action. Likewise, there will be cases (as with a parent and a child in a dangerous situation) where both anxiety and compulsive checking will be appropriate. In all these cases, then, we can see reason for the coexistence of multiple emotional states and for positive feedback between them (for example, the more anxiety-provoking the situation, the greater should be the level of compulsive checking). It is this tendency towards functional coexistence of several adaptive emotional states which, we believe, provides part of the basis for the comorbidity of the different anxiety disorders. The biasing of all these systems by a common underlying personality variable (Chapter 12) provides another part.

These considerations elucidate the difficulties faced in the diagnosis of anxiety disorders: a specific symptom can arise either because of pathology of its specific controlling centre, or because of changed activity in that centre secondary to pathology in some other part of the distributed defence control network. Given that, at least in human beings, display rules can greatly modify the external signs of emotion and that conditioning can alter the internal signs, it follows that pathology in a specific control centre may give rise to a cluster of symptoms, and that those over which the given centre has the most direct control will not necessarily even be the most salient. We consider these possibilities in more detail in relation to individual symptomatological and diagnostic categories below.

A final necessary distinction, before moving on to these diagnostic categories, is that between anxiety itself and the anxiety-related disorders. To a large extent, we argue below that pathological anxiety is a specific delimited condition to be distinguished from pathological panic, phobia, and obsession. On the other hand, we see it as reasonable to group these latter conditions together as 'anxiety-related disorders' since (i) the symptoms show a high level of co-occurrence, and the syndromes high levels of comorbidity; and (ii) in the cases of panic and obsession comorbid anxiety is a significant feature of their presentation in the clinic.

Let us now consider the individual disorders in more detail. We take them in order of increasing defensive distance (Chapter 2), and deal with them for convenience of exposition under the headings used by DSM-IIIR. Inevitably, there will be much speculation in our attempts to relate the symptoms of these disorders to activity in the brain. However, the rapid developments in methods for imaging the human brain in action make it particularly valuable, albeit also hazardous, to record these speculations now, since in all probability it will shortly be possible to test them experimentally.

11.12 PANIC DISORDER

Following Graeff and others (Chapter 6) we identify panic attacks *per se* with activity in the periaqueductal grey, although the cognitions reported in relation to an attack must be mediated cortically, and these form a significant component of panic disorder as this presents in the psychiatric clinic. What conclusions does this identification allow us to draw about panic as a symptom and about panic disorder?

Consider, first, panic as a symptom. Panic attacks should occur in normal individuals under normal ecological circumstances when defensive distance is small and the fear stimulus intense. Given that defensive aggression is a relevant strategy with small defensive distance, the increased hostility associated with clinical panic is consistent with this analysis (Fava *et al.* 1993). It follows from the arguments adduced in the previous section that panic attacks should in addition occur as symptoms when there is no abnormality of the panic control system itself, but this receives abnormal input as a result of activity elsewhere. These arguments have supposed that many components of the defensive system can be activated concurrently, and that activity in one area can feed into another. Extremely high levels of fear (amygdala) or anxiety (septo-hippocampal system + amygdala) could, then, activate the periaqueductal grey, releasing it from the serotonergic inhibition that would normally keep it in check. Thus, as suggested by Marks (1987), some cases presenting with panic attacks will simply be cases of extreme anxiety. The parallel case in which extreme fear (i.e. avoidance without a conflicting approach tendency) might result in panic would not often present in the clinic since, under normal circumstances, the pathologically phobic patient will succeed in avoiding the fear stimulus and so not suffer from panic attacks.

Now let us consider panic as a disorder. From a neurological point of view this should result from dysfunction of the periaqueductal grey itself, resulting in spontaneous panic attacks in the absence of any specific fear- or anxiety-provoking stimulus. Given this basis for panic disorder, it is theoretically possible for mild panic attacks to occur, even though 'mild panic' sounds like a contradiction in terms. In this, we are close to the view of Marks (1988b, p. 127) that a deficiency in DSM-IIIR is that it does not allow for the possibility that 'panics need not be so severe or abrupt in onset as to be called "panic" at the start'. One can imagine two ways in which panic disorder could arise. First would be the production of spontaneous activity in the periaqueductal grey via, for example, seizure. This would be expected to occur extremely seldom and can probably be ignored for normal clinical purposes (but see Keck *et al.* 1993; and also Almayehu *et al.* 1995 on the role of parietal lobe seizures). Second would be an abnormal sensitivity of this structure to its normal adequate stimuli, or an abnormal input to it as a result of the occurrence of these stimuli. This would manifest itself as a decreased threshold for panic attacks in a provoking situation. Under normal conditions in human society there will be few appropriate external stimuli to demonstrate this sensitivity in any selective fashion. However, there is one internal stimulus, suffocation, which is both an adequate stimulus for activation of the periaqueductal grey and likely to occur, with moderate intensity, under everyday conditions. This, therefore, could give rise to apparently spontaneous panic attacks.

Klein (see for example 1995 for a brief review) 'proposed the hypothesis that a panic attack may be a suffocation false alarm, that is, patients have a pathologically lowered threshold for estimating suffocation, which would account for the dyspnoea [seen in panic disorder]' (Klein 1995, p. 62s; see also Papp *et al.* 1993). The distinction between panic-as-symptom and panic-as-disorder, noted earlier, is able to bring order at this point into what otherwise might be rather confusing data on the relative frequency of occurrence of dyspnoea in different clinical conditions. Given Klein's hypothesis, dyspnoea would be expected in particular to be a characteristic of patients with a primary diagnosis of

panic disorder; and, as Klein notes, this is indeed the case. On the other hand, dyspnoea would not be expected to occur in other forms of anxiety-related disorder; and this is so for both phobic and obsessional patients (Asmundson and Stein 1994). Nor would dyspnoea be expected to occur in panic provoked by strong environmental stimulation. This again is the case: 'in intense, fearful situations, such as those experienced and recounted by World War II veterans, symptoms of fear include palpitations, trembling, and sweating, but not dyspnoea' (Klein 1995, p. 62s). Klein's suffocation hypothesis can also account for the occurrence of panic during dreamless sleep (e.g. Rosenfeld and Furman 1994); for the cardiac and respiratory abnormalities of patients with panic disorder during sleep (Koenigsberg *et al.* 1994); for the ease with which respiratory challenges elicit panic; and for the fact that patients with panic disorder terminate breath holding earlier than do social phobics and controls (Asmundson and Stein 1994). There is also evidence that panic disorder may involve a more pervasive problem with autonomic control (Middleton *et al.* 1994); thus, some panic patients may be characterized by physical reactions other than dyspnoea.

If we consider the history of panic disorder (Holt 1990, p. 55), we find that 'during World War I, the symptom complex of palpitation, breathlessness, chest pain, fatigue and digestive disturbance originally identified by Hartshorne (1864) was studied intensively, being called Irritable Heart Syndrome, Disordered Action of the Heart or Da Costa's Syndrome. . . . However, no evidence of heart disease or heart abnormalities could be found. . . . Lewis (1919) preferred the term Effort Syndrome . . . "because these symptoms and signs are largely, in some cases wholly, the exaggerated physiological response to exercise" . . . [but] 76% [of Effort Syndrome patients] reported the involvement of fear in their condition, 70% had attacks at rest (i.e. not after exercise), 50% were depressed and 70% showed evidence that they were anxious.' Thus it appears that only very recently has what is now termed panic disorder been seen as a particularly psychological condition. It seems possible, also, that there has been some drift in the selection of patients from a group of whom only 70% were anxious to one in which 100% are anxious. Certainly, even very recently 'there has been little consistency in measures or even types of assessment used to characterize and follow up [panic] patients' (Shear and Maser 1994, p. 346).

We suggest that pathological panic can occur in any of the following ways (these being poles of dimensions rather than exclusive categories).

First, there will be people who are abnormally sensitive to panic stimuli, but who are at the opposite, stable extrovert, pole to the neurotic introverted personality that predisposes to anxiety. In Shear and Maser's (1994, p. 349) words, such patients experience a limited 'physiological, form of panic attack and will have, in our view, the purest form of panic disorder. However, they will not usually present in the psychiatric clinic because the physiological component of the panic attack is not exacerbated by a cognitively-mediated reaction from the defence system—and indeed do not fit with the conventional view that "anticipatory anxiety is a central symptom of panic disorder".' Aoki *et al.* (1994, p. 56), for example, report that in a non-patient population 'panickers seeking help showed a tendency to be less extroverted than those not seeking medical care'. As a result, stable extroverts may either ignore the attack, or seek advice from a heart specialist (who will tell them there is nothing wrong with their heart), and the fear and/or anxiety components of the attack (already minimal) will extinguish.

Second, we have individuals whose sensitivity to panic stimuli is normal but who have a concurrent anxiety or phobic disorder and who, because this is extreme, experience an increased frequency of panic attacks. These will not normally be diagnosed as suffering from panic disorder.

Third, there will be those who have both a high sensitivity to panic stimuli and a highly neurotic introvert personality. These are likely to have a concurrent anxiety or phobic disorder, in which case they should be diagnosed as having panic disorder comorbid with the anxiety or phobic disorder.

Fourth, there will be people who are particularly sensitive to panic stimuli, but who are of an intermediate personality type (neither neurotic introvert nor stable extrovert). Their panic attacks and panic disorder will be full-blown, because the physiological panic attack (which may normally be very brief, since high pCO_2 is easily corrected) triggers a reaction from the defence system engendering fear (and on later occasions anxiety also) which immediately feeds back and potentiates the physiological panic attack. Here, we depart slightly from Klein's views, since we see this psychological component of the attack as driven by the fact that the periaqueductal grey is supersensitive not only to suffocation stimuli but also to fear and anxiety stimuli (see for example Abelson and Nesse 1994). Thus the respiratory and cardiac aspects of panic disorder are most evident, in our view, because they are the usual trigger for an attack, especially early in the development of the disorder when high levels of anxiety are relatively rare. However, when anxiety is itself elevated, this can either increase the sensitivity to panic stimuli yet further, or actually generate a panic attack. This account is consistent with the observation that patients treated with extensive psychological therapy 'were essentially free of panic attacks during the follow up period [of 5 years and] the most significant change, in their view, was that, if an attack did occur, they knew how to deal with it and did not overreact to it, or let it limit their mobility' (Franklin 1990, p. 89).

We can speculate that in this group of patients prior to the clinical phase of their disorder, there was an initial low level of panic determined solely by physiological abnormality, and that after treatment their level of panic returned to this low level. The clinical phase of their disorder, on this view, resulted from the conditioning of anxiety to environmental stimuli by the spontaneous panic attacks and hence chronic anxiety. This anxiety could increase the frequency of panic attacks through two routes, both independent of the behavioural inhibition component of anxiety and so of putative hippocampal involvement. First, chronic anxiety is characterized by chronic autonomic arousal; this should increase the probability of panic attacks, given the known capacity of adrenaline infusion to elicit panic (e.g. Veltman *et al.* 1996; see also Lader and Tyrer 1975). Such a conditioned increase in arousal is formally equivalent to fear-potentiated startle, and so is likely to be generated by a direct threat input to the amygdala. Second, the detection of threat itself could activate the periaqueductal grey and produce panic directly, since even the highest level of the defence system, represented by the cingulate cortex, has a direct input to the periaqueductal grey (see Appendix 2). Overlaid on this long-term, conditioning-based vicious spiral of exacerbation of panic by anxiety, and of anxiety by panic, there could be an acute suppression of panic involving the behavioural inhibition component of anxiety. It is via this latter, antagonistic relationship (see Graeff 1994) that one can account for the phenomenon of relaxation-induced panic, sometimes

to the degree that panic regularly wakes the sufferer from sleep; although Graeff's account requires the additional, but not implausible, assumption that the descending inhibitory controls from the forebrain are switched off, as relaxation or sleep sets in, prior to the switching off of the lower brain systems that mediate panic symptoms. Given this account, relaxation-induced panic should occur only in patients with primary panic disorder (as defined by, say, CO_2 challenge), not in those whose panic attacks are a consequence of primary anxiety.

This type of analysis can fairly easily accommodate the fact that, in patients referred for chest pain who have no sign of cardiac malfunction, 60% can be then diagnosed with panic disorder (Carter *et al.* 1994). We may speculate that this group probably suffers from primary panic disorder (in our terms), with about 40% of them being stable extroverts and 60% in the middle of the relevant personality continua. That is, neurotic introverts with panic would have been initially diagnosed as panic disorder, and so would not enter the cardiac sample; those with intermediate personality type would have an insufficient level of anxiety for a primary diagnosis of panic disorder, and so would enter the cardiac sample and subsequently be diagnosed with panic disorder; and those with stable extrovert personality types would have an insufficient level of anxiety to be diagnosed as panic disorder using the currently accepted criteria.

As well as the temporary, functional release from forebrain inhibition of lower levels of the defence system which, as noted above, appears to underlie relaxation-induced panic, there is also evidence of chronic loss of such inhibitory control due to structural damage in the brain, as assessed by magnetic resonance imaging. Consistent with our overall model, this damage has been localized to the septo-hippocampal system. Thus, Dantendorfer *et al.* (1996) first subdivided an initial sample of 120 patients with panic disorder into those with (29%) or without abnormalities (but not sufficient to constitute epilepsy) in conventional scalp electroencephalographic recordings, and then further examined the group showing such abnormalities for structural brain damage. A remarkably high proportion (61%) of this group turned out to have such damage, most frequently affecting the hippocampus and/or septal area. (As we shall see, hippocampal damage has been noted also in post-traumatic stress disorder.) The aetiology of such damage is undoubtedly heterogeneous. Two likely pathways, however, are perinatal anoxia (the hippocampus being particularly vulnerable to anoxic insult) and degeneration in hippocampal pyramidal cells due to excessive exposure to corticosteroids (McEwen *et al.* 1992). Given the elevated levels of corticosteroid hormones that occur during prolonged exposure to stress (e.g. Sapolsky *et al.* 1990), the latter pathway is likely to participate in a second type of vicious spiral, one on a longer time base than the immediate interactions between anxiety and panic considered above. In this slower spiral, an initial period of prolonged exposure to stress would involve heightened activity in lower levels of the defence system, coupled with inhibition of their output involving the septo-hippocampal system; psychologically, this period would be expected to be characterized by combined symptoms of panic and anxiety. Subsequently, after steroid-induced damage to the hippocampus, one would expect the symptomatology to be dominated by panic. In addition, the hippocampal damage should give rise to appropriate cognitive deficits. In agreement with the latter prediction, Maierhofer *et al.* (1998) have shown that patients with panic disorder show deficits in conditional, but not simple, discrimination learning

of the same kind that Daum *et al.* (1991) had previously observed in patients who had undergone temporal lobe resection for the treatment of epilepsy.

Given this general analysis of panic, anxiety, and their interactions, and provided there is appropriate differentiation of receptor types, it should be possible to produce drugs that are pure anti-panic agents and others that are pure anxiolytic agents. Some anti-panic drugs would be expected to reduce panic quite generally (by acting on cellular activity in the periaqueductal grey); others might reduce panic only in a subset of cases of panic disorder (by acting, for example, solely on the terminals which carry the afferent anoxic signal to the periaqueductal grey), leaving panic attacks undisturbed in cases where panic is a symptom of anxiety or phobic disorder. Pure anxiolytic drugs would be expected to have no effect on the purest cases of panic disorder, to have an ameliorative effect in panic disorder where this is exacerbated by anxiety, and to be fully effective in those cases where a modest incidence of panic attacks is consequent on very high levels of anxiety. There do indeed appear to be drugs which act differentially on anxiety and panic, justifying for that reason alone a view of panic as separate from anxiety. The same appears to be true of panic and depression (see for example Rosenberg and Jensen 1994). However, in this respect our analysis may be too conservative in that we imply that there will often be a tight linkage between anxiety and panic. Buspirone has been reported to be ineffective in panic disorder (Sheehan *et al.* 1990). On our analysis it should have had at least some ameliorative effect in at least some cases. There are two possibilities here. First, patients selected for test with buspirone may have been of a more purely panic type rather than mixed panic–anxiety. Second, and more likely on our theory, the paradoxical actions of buspirone on 5HT systems or on cortisol levels (Appendix 1) may have produced a counterbalancing direct exacerbation of panic, offsetting any indirect amelioration produced by its anxiolytic action.

In all of the above we have talked glibly about the relevant inputs to the periaqueductal grey as if they were simple internal or external stimuli. But, of course, it is inherent in the idea of defensive distance that the stimulus must be interpreted as threatening for the reaction to be obtained. Rats and other animals, for example, can learn that that 'innate predator', a human care-giver, is non-threatening. At this point, therefore, we must link our theory with recent ideas on the role of cognition in panic and panic disorder. This is especially so, since there is a tendency for people, mistakenly, to contrast biological and cognitive explanations of behaviour.

A succinct review of one cognitive theory of panic is provided by Clark and Ehlers (1993; see also Clark 1986). This theory has many similarities with the analysis we have provided above. In particular, the main generator of the panic attack is presumed to be a positive feedback between bodily sensations and perceived threat. In the theory 'it is assumed that the crucial event is a misinterpretation of certain bodily sensations' (Clark and Ehlers 1993, p. 132). Such misinterpretation may result from an increased sensitivity of perception of some specific sensations such as heart rate (Näring and van der Staak 1995; but see Barsky *et al.* 1994). In favour of the theory is the fact that panic patients 'were significantly more likely to choose negative interpretations of the key bodily sensations . . . but were no more likely to choose negative interpretations of the other ambiguous events' (Clark and Ehlers 1993, p. 133); the fact that cognitive therapy appears particularly effective with panic disorder (for various opposing views see Clark

et al. 1994; Klein 1994; Marks *et al*. 1994 and reply by Clark *et al*.; Shear *et al*. 1994; and also our own review in Chapter 13); the fact that appropriate instructions before biological challenges reduce the probability of panic; and the fact that heart-beat discriminability predicts the maintenance of panic attacks (Ehlers and Breuer 1996).

Differences between our analysis of panic and this cognitive theory occur at two points. First, the cognitive theory is intended to include only a subset of the phenomena we have included. Second, there are consequent terminological differences in the uses of words such as 'panic' and 'panic disorder'. We included in our analysis of panic disorder cases of, for example, 'benign heart pain', arguing that in terms of the core biological component of panic, these cases are not essentially different from the 60% of the same heart pain population who received a subsequent diagnosis of panic. These patients are excluded from the cognitive theory (and, of course, conventional diagnosis of panic disorder), but they are germane for the following reasons. 'Surveys suggest that 7–28% of the normal population will experience an occasional unexpected panic attack. . . . However, the cognitive theory assumes that an individual only goes on to develop the rarer condition of repeated panic attacks and panic disorder . . . if he or she develops a tendency to interpret these perceived autonomic events in a catastrophic fashion' (Clark and Ehlers 1993, p. 132). This 'tendency to interpret' is equivalent to the cognitive component of the conditioning we have postulated as occurring in a person with a neurotic introvert personality. However, if personality type were all there were to it, we would expect generalized anxiety disorder with occasional panic attacks to progress rapidly to full-blown panic disorder; and this does not appear generally to be the case. The cognitive theory can accommodate this by postulating a specific *additional* personality factor relating to the perception of somatic events. But the need for an additional explanatory factor when excluding some of the available data is not parsimonious.

11.13 SPECIFIC PHOBIA

Specific phobia (which, consistent with DSM-IIIR, we distinguish from agoraphobia and social phobia) involves the active avoidance of a specific phobic object. Thus, on the theory presented in this book, specific phobia represents fear essentially unmixed with anxiety and so should be insensitive to anxiolytic drugs in the same way as the items in the Blanchards' fear/defence test battery (Chapters 2 and 4). This is indeed the case (Sartory *et al*. 1990). The brain activity elicited by phobic objects is also insensitive to anxiolytic drugs (Fredrikson *et al*. 1995). On both psychological and pharmacological grounds, therefore, we must distinguish specific phobia from anxiety. It is proper, nonetheless, that we should discuss specific phobias in the context of the anxiety disorders, since the presence of a phobic disorder should be capable in some cases, through normal conditioning processes, of giving rise to anticipatory anxiety. (The converse, however—anxiety, as in generalized anxiety disorder, leading on to specific phobia—theoretically should be, and apparently is, rare.)

If specific phobia is not itself the result of anxiety disorder (whatever its capacity to elicit secondary anxiety), what is its aetiology? Previous theories of phobia, derived from the animal laboratory, have emphasized conditioned fear as a model for phobias with

few and clearly delineated precipitating stimuli (H. J. Eysenck and Rachman 1965; Gray 1971); and, consistent with this view, behaviour therapy has been particularly successful in its treatment (e.g. Mathews 1978). Although such phobias are of relatively little clinical importance (Marks 1969), these theories could apparently account for them in terms of a well-defined experimental model, giving them considerable plausibility.

Before discussing the conditioning theories in greater detail, let us look at phobia in the context of our model. First, following the Blanchards' analysis (Chapter 2), we would expect there to be both innate and conditioned fear stimuli to which an excessive neural reaction could produce a maladaptive response. Second, following LeDoux's analysis (Chapter 6), we would expect conditioned fear stimuli to have been created by strengthening of synapses in the amygdala as a result of the conjunction of input to this region from lower and higher sensory-processing areas with input from an appropriate unconditioned stimulus such as pain. A particularly simple explanation of the evolution and existence of innate fear stimuli would then be the selection of those members of a population in which conditioning to stimuli emanating from certain common sources of danger was particularly easy. A wide range of stimuli have access (via conditioning) to the fear mechanisms located in the amygdala. Provided random mutation (or the shuffling of multiple genes during sexual reproduction) continually maintains the variance in such a population despite a shift in its mean, the end-point of such an evolutionary process will be animals demonstrating 'zero trial learning': that is the production, as an unconditioned response, of what in previous generations was a conditioned response. The reverse process can also occur. A key technical feature of the Blanchards' experiments was the use of wild rats, for which the experimenter is an innate predatory stimulus, as compared to laboratory rats, which have largely lost this tendency as the result of selective breeding.

Given this view of conventional fears what would we expect of pathological phobias? First, there could be a supersensitivity of the amygdala to fear conditioning (perhaps as the result of an overreactive reinforcement signal) or excessive output from the amygdala after fear conditioning has occurred. This should give us a patient who reacts with excessive fear to a large number of dangerous objects. Both conditioned and innate fear stimuli should be affected in this case. Second, the natural variation in the population (or possibly some special feature of ontogeny) would produce a greater than normal reaction to innate fear stimuli. This should produce results more selective to innate fears and, most likely, selective to specific fears in specific individuals.

The first of these two possibilities is, in essence, the basis for the 'standard' conditioning theory of phobias, usually attributed in the first instance to Watson and Rayner (1920). This holds that there are a few stimuli that are innately capable of eliciting fear reactions in young human beings (Watson listed loud noise, pain, and sudden loss of support). The strange panoply of adult human phobias is then thought to arise through Pavlovian conditioning between these innate stimuli (as UCSs) and a random assortment of conditioned stimuli that happen to achieve the right temporo-spatial association with them.

Some case reports of adult phobics confirm Watson's findings that phobias can be learned by association. The most interesting of these were described by Bagby in 1922. His first case was that

of a young woman who had retained since childhood a severe fear of running water. . . . As a little girl of seven she had [become trapped alone] . . . wedged between two rocks, with a waterfall noisily pouring over her head. . . . Although [this] and similar cases provide evidence that phobias can be acquired through the simple pairing of an alarming occurrence and a hitherto neutral situation, the sad fact [for conditioning theories] remains that such stories are rare. Few phobics can recall the precipitating incident for their fear. Furthermore, evidence from other sources raises problems that this theory, which claims that phobia is learned by association, cannot resolve . . .

War presents us with the conditions for a large-scale experiment when great numbers of civilians are exposed to bombing attacks. In such attacks loud noise is paired with a variety of hitherto neutral events. According to the theory of learning by association, this type of conditioning should generate a large number of phobias. Yet . . . the vast majority of children, 96 percent, were unaffected. . . . This, then, is a failure of the conditioning theory of phobia on a gigantic scale. (Agras 1985, pp. 45–7.)

Not only is there this negative evidence against the conditioning theory, but there are positive objections also. H. J. Eysenck (1979) objected to the theory on a number of grounds. The most important are the following. (i) The stimuli that elicit phobias are not a random sample of stimuli: some (e.g. closed spaces, snakes, spiders) are greatly over-represented; others (e.g. cars) that are associated with objective dangers are under-represented. (ii) Phobic stimuli unaccompanied by their UCSs ought to undergo extinction; of course, they do not, or phobias would not constitute a psychiatric problem. To these can be added a third objection (Gray 1979a). (iii) The times of onset of phobias are not a random sample of ages; there is a predominance of onsets in early adult life (Marks 1969).

The first of Eysenck's objections is well founded, but the second less so. Most well-learned responses are habitual and should persist unless conditions change (McNaughton 1989, pp. 124–8). In the specific case of avoidance learning, there are reports in the animal literature of an essential lack of extinction in the total absence of the reinforcer (Gray 1987, Chapter 11). On this view, the rapid extinction of *appetitively* motivated responses is an aberration which results, not because the response is no longer reinforced as such, but because '*omission* of an expected reward . . . is a highly salient event which can actively suppress responding' (McNaughton 1989, p. 127, our emphasis; see also Appendix 7). Strong resistance to extinction is not seen, however, in classical as opposed to instrumental fear conditioning. 'Fears conditioned by association tend to be short-lived. Most conditioned reflexes will weaken and disappear after a few exposures to the fear-provoking event. . . . [However,] avoidance of a feared situation blocks the normal process of unlearning the fear response. . . . In the case of air raids, [mentioned above,] which are unavoidable for a civilian, we would expect the fear response to be transient. *Only if the phobic situation is avoidable, as it usually is, would we expect the phobia to be long-lasting*' (Agras 1985, p. 47). Consistent with these arguments, the key therapeutic factor in the treatment of phobia is the total amount of time for which the patient is exposed to the phobic stimulus—the greater the exposure, the greater the therapeutic effect (Gelder *et al.* 1973; Marks 1973; Levis and Hare 1977; Teasdale 1977). (This issue is discussed in greater detail in Chapter 13, Section 13.1.) Other factors such as attention do have some influence, but their precise role is still unclear (Marks and

Marks 1990). Lader (Lader and Wing 1966; Lader and Mathews 1968) and later Watts (1971, 1979) suggested that the common element in quite different methods of behaviour therapy (e.g. systematic desensitization, flooding) is that they allow responses to the phobic stimulus to habituate. This is what we would expect from the analysis of avoidance conditioning.

There remains the issue of the non-random nature of phobic stimuli, Eysenck's objection (i) above. In an effort to save the conditioning theory from both this and objection (ii), Eysenck (1979) made use of Seligman's (1971) concept of 'preparedness' and allied it to a new theory of his own. According to this theory some stimuli are 'prepared' (by Darwinian evolution) to enter into an association with aversive UCSs. (This is different from the view of innate fear stimuli elaborated below.) Furthermore, once such an association has been formed, the phobic power of the prepared CS, far from extinguishing, can be further increased by presentation without a following UCS (a process termed 'incubation'). However, while it is an improvement on the simple conditioning theory, Eysenck's alternative is also unsatisfactory (Gray 1979a). First, it does not meet objection (iii) above. Second, the experimental evidence for incubation is poor (Bersh 1980); and, as we noted, there is in any case no need to invoke any special mechanism to account for slow extinction. Third, the concept of preparedness for conditioning, in this simple form, is unacceptable. This last point requires justification; especially since there are data, particularly from Öhman's group in Uppsala, which apparently offer strong support both for the concept of preparedness and for its relevance to the genesis of phobias. Close inspection of these data shows, however, that they are not quite what they seem.

11.13.1 **Öhman's experiments**

The main experimental paradigm used by Öhman and his collaborators was that of differential classical conditioning of autonomic responses (usually the skin conductance response) with an aversive UCS (usually a mild electric shock) and visual stimuli (presented as slides) as CSs. The experiments were conducted with normal human subjects, most often undergraduates. The key innovation lies in the choice of CSs. These consist of objects which are either commonly (e.g. snakes, spiders: 'fear-relevant stimuli') or rarely (e.g. flowers, mushrooms: 'fear-irrelevant stimuli') found as the objects of clinically presenting phobias.

Let us first present Öhman's findings as seen from the perspective of the preparedness hypothesis (Seligman 1971; Öhman 1979). According to this view, phobias develop when prepared stimuli are followed by an aversive UCS. The resulting conditioned fear response is said to be formed very rapidly (indeed in one trial), to be extremely resistant to extinction, and to be resistant to control by rational argument. From this it follows that Öhman's fear-relevant stimuli should be more potent CSs for the skin conductance response, provided an aversive UCS is used, than his fear-irrelevant stimuli. The CR to fear-relevant stimuli should be formed more readily, it should be more resistant to extinction, and it should be less open to cognitive control.

With the exception of speed of acquisition, these predictions are in general confirmed by Öhman's results. When a shock UCS is used, the differential conditioned skin

conductance response (that is, the difference in the magnitude between CS+ and CS–) is more resistant to extinction if the stimuli are fear-relevant rather than fear-irrelevant (e.g. Öhman *et al.* 1976). Furthermore, if at the start of extinction instructions are given that there will be no more shocks, the differential conditioning disappears at once with fear-irrelevant stimuli, but only slowly with fear-relevant stimuli (Hugdahl and Öhman 1977). By contrast, with stimuli which are fear-relevant, but which could *not* be prepared by evolution to be so, there was no such effect (Hodes *et al.* 1977, cited by Öhman 1979).

This pattern of results seems to offer strong support for the preparedness hypothesis. But closer examination reveals several curious aspects of the findings which fit ill with the central assertion of preparedness theory, namely, that prepared (fear-relevant) stimuli are more readily *associated* with aversive UCSs.

First, most of the data has been obtained with the skin conductance response. It is well known that this response participates in the orienting reflex (Sokolov 1960). Although the orienting reflex can be conditioned, it is not clear that this depends on the same processes as do other forms of Pavlovian conditioning; and, in any case, the preparedness theory is concerned with conditioned fear, not conditioned orienting. Öhman *et al.* (1978) have attempted to meet this objection by arguing that the skin conductance change observed in response to a fear-relevant CS for a shock UCS is a conditioned defensive reaction (Sokolov 1960). Their evidence consists in certain differences between the skin conductance response measured on the dorsal and palmar surfaces of the hand respectively; but the data they report are complex and by no means all in agreement with the conclusion they reach. More importantly, an attempt to substantiate the same conclusion by measuring heart rate in conjunction with the skin conductance response was unsuccessful (Fredrikson and Öhman 1979).

Other features of Öhman's data also suggest that the response he measures is directed at the CS (as is a conditioned orienting response) rather than being anticipatory of the UCS (as is commonly the case in classical conditioning; Pavlov 1927). In most of his experiments the CS is presented for 8 seconds and the CR is measured separately during the first and second halves of this interval. It is only the response measured during the first half of the CS period which behaves according to the predictions of preparedness theory; the response measured in the second half of this period is unaffected by the major experimental manipulations (e.g. Öhman *et al.* 1976). This is quite unlike a normal CR, which is usually maximal just prior to UCS onset (Pavlov 1927). Furthermore, Öhman (1971) has himself presented evidence that the skin conductance response measured during the first 4 seconds of the CS period behaves like an orienting response. Thus the most plausible interpretation of the findings in which a fear-relevant CS is paired with a shock UCS is that this *causes particular attention to be paid to the CS*.

Now this conclusion might be acceptable to proponents of the preparedness theory, since it could be said that we have done no more than redescribe the conditioned response: rather than conditioned fear, it is now conditioned attention to the CS. But there are also other features of Öhman's results which are inconsistent with the notion that prepared stimuli are more ready to enter into an association with aversive UCSs, no matter how we then describe the conditioned response.

If the same processes are involved as in normal conditioning, it is in the first place surprising that the skin conductance response elicited by fear-relevant stimuli is insensitive

to several experimental parameters which affect other forms of conditioning quite powerfully: number of acquisition trials (Öhman *et al.* 1975), the difference between trace and delay procedures, and the CS–UCS interval (Hugdahl and Öhman 1980). In rebuttal of this point, proponents of preparedness theory might claim that, although prepared stimuli undergo conditioning, it is a special kind of conditioning (Seligman 1971). The appeal to a special kind of conditioning goes against the current trend to see all plasticity as dependent on either simple association (based on long-term potentiation) or simple reinforcement, based on activity-dependent facilitation (Kandel and Hawkins 1992). However, we can adduce a stronger objection to the idea that any type of stimulus preparedness (in Seligman's sense) is involved at all.

According to the preparedness theory, fear-relevant stimuli such as snakes and spiders are particularly ready to enter into a positive association with a negative reinforcer. From this we would predict that such a stimulus used as a CS+ will form a swift attachment to the UCS. However, we would predict that the same class of stimulus will also readily form an attachment to the UCS when it is used as a CS–. Thus in a within-subjects differential conditioning experiment, with both the CS+ and the CS– being fear-relevant stimuli in one group of subjects and fear-irrelevant in a second group, we would predict that the elicitation of conditioned responses would occur much earlier to fear-relevant than fear-irrelevant stimuli, but that differential conditioning between the CS+ and CS– would be unchanged or might even be slower with fear-relevant stimuli. Yet, it is precisely this flawed within-subjects differential conditioning paradigm which has provided the major apparent support for the preparedness hypothesis (e.g. Öhman *et al.* 1976). But the preparedness theory has no reason to predict that it should be easier to *discriminate* between a pair of prepared stimuli than between a pair of unprepared stimuli.

That something other than preparedness for conditioning to aversive stimuli is going on in these experiments is indicated by the greater skin conductance response to fear-relevant than to fear-irrelevant stimuli even when *no association with an aversive UCS is involved at all.*

To begin with, fear-relevant stimuli elicit a greater skin conductance response when simply presented repeatedly in a habituation paradigm (e.g. Öhman *et al.* 1974; Hugdahl *et al.* 1977). Furthermore, if a fear-relevant stimulus follows another such stimulus in a habituation paradigm, the skin conductance response is potentiated, a phenomenon not seen with fear-irrelevant stimuli (Hygge and Öhman 1978). More directly related to shock conditioning experiments is the observation (Öhman *et al.* 1974) that the difference between the unconditioned skin conductance responses to fear-relevant and fear-irrelevant stimuli, respectively, is considerably magnified if the subject is given experience with a mild electric shock before the stimuli are presented and is told simply that shock will be delivered during the experiment. In the light of this observation it is not surprising that the response to fear-relevant stimuli is particularly prone to sensitization when shocks are in fact delivered, but in a random relationship to the stimuli (Öhman *et al.* 1975). In this experiment separate groups of subjects were exposed to pictures of either snakes or houses in a stimulus-only condition, to pictures and shocks randomly intermixed (truly random control), or to pictures predicting shocks (classical conditioning). During acquisition, the skin conductance response in the first half of the CS period was significantly larger for snakes than houses in the truly random and classical

conditioning groups, but not in the stimulus-only condition; furthermore, the truly random and classical conditioning groups did not differ from each other.

These results strongly suggest that the other phenomena observed by Öhman's group are largely due to sensitization and have nothing at all to do with conditioning as this is usually understood, especially since the experiment described in the previous paragraph is apparently the only time they have controlled for this possibility. However, in the extinction phase of the experiment, the pattern of results was consistent with a true conditioning effect specific to the fear-relevant stimulus: the response was significantly greater in the fear-relevant conditioning group than in the fear-relevant truly random controls, and significantly greater in the fear-relevant than in the fear-irrelevant group (Öhman *et al.* 1975). Thus it appears that fear-relevant stimuli elicit greater skin conductance responses initially, and that their capacity to elicit such responses is differentially potentiated by shock or the threat of shock, even though this is not specifically associated with them; but that this capacity is *also* differentially potentiated by associative conditioning, at least when measures are taken during extinction. But preparedness theory predicts only the last of these findings.

Given the evidence that fear-relevant stimuli elicit a degree of emotional excitement in the absence of any specific conditioning, even this last finding can be explained in a manner that does not call upon the concept of preparedness. For it has been shown (in experiments in which preparedness appears to play no part) that conditioning is facilitated by similarity between the events to be associated (Rescorla 1978). If the capacity to elicit fear (or a 'lurking' fear, in Valentine's (1930) terms; see Gray 1987b, Chapter 2) provides a dimension along which such similarity can act, this principle could account for the greater ease of association between two innately fear-eliciting stimuli, be they two pictures of potentially phobic objects (snakes, spiders), or one such picture and an electric shock. This explanation can also account for the great ease with which, in Mineka's (1987) experiments, neonatal rhesus monkeys formed an association between video clips of snakes and fearful adult monkey models, respectively. Thus there is no need to call upon preparedness, as proposed by Seligman (1971) and Eysenck (1979), to account even for the extra associability of such pairs of stimuli.

Öhman has recently shown, using a masking paradigm (Öhman and Soares 1998), that both the low-level elicitation of fear by fear-relevant stimuli and the conditioning of autonomic responses to them can occur independently of conscious awareness and then affect subsequent cognitive processing. He links this to LeDoux's idea of parallel subcortical and cortical circuits controlling fear conditioning (Chapter 6) and the idea

that activity in the former could be used as input to the latter. In LeDoux's (1996) words, 'the feeling of being afraid results when we become consciously aware that an emotion system in the brain, like the defense system, is active' (p. 268). Because the fear system can be nonconsciously activated, which may bias the significance evaluation of the stimulus, the emotional system would prompt the cognitive system to expect bad things to happen. . . . For example, anxiety patients have a generalized expectancy bias to become personally destined to experience aversive outcomes. . . . [The] tendency to judge aversive outcomes as somewhat likely after fear-relevant stimuli is considerably enhanced when the context is made aversive . . . [and] illusory correlations between fear-relevant stimuli and aversive events . . . require that the fear-relevant stimulus actually generates some fear. It appears, therefore, that the nonconsciously activated fear system alerts the cognitive

system evaluating threats . . . to expect aversive events. . . . In most instances, this is an adaptive mechanism . . . what Mineka (1992) called adaptive conservatism, that is, a tendency to be cautious and play it safe when fear is activated. However, if the emotion persists, a vicious circle may be entered, in which anxiety produces biases to expect and discover further threats, which further enhance anxiety, and so on (Öhman and Soares 1998, pp. 80–81.)

11.13.2 **Phobic stimuli as innate fear stimuli**

In summary of the previous sections, it seems likely that a small proportion of phobias can be established via conventional conditioning. Indeed, modern conditioning theories and the data upon which they are based have tended to extend the range over which classical conditioning might be operative in the formation of phobic reactions, especially when one takes into account the effects of such phenomena as latent inhibition or UCS revaluation (Davey 1992; Mineka and Zinbarg 1996). However, in the majority of cases the over-representation of a relatively small set of phobic objects suggests that some other, or additional, explanation is required. Öhman's results show that these same 'fear-relevant' stimuli come to elicit more persistent skin conductance responses in consequence of an association with an aversive UCS. Given the lack of a similar effect when fear-relevant but phylogenetically novel stimuli (rifles and revolvers; Öhman 1979) are used as CSs, it is justifiable to replace the term 'fear-relevant' by a term which implies some innate contribution to the observed phenomena. However, the classical concept of preparedness does not fit with the bulk of Öhman's results. This concept implies that the special characteristic of prepared stimuli is that they can more readily become conditioned fear stimuli as the result of a Pavlovian conditioning process. This hypothesis does not explain: the fact that the main skin conductance effects occur in the first half of the CS period; the special affinity of prepared stimuli for the skin conductance response as opposed to other response measures; the fact that prepared stimuli elicit strong skin conductance changes in the absence of any conditioning; that they are susceptible to sensitization; that they are particularly easily differentiated into positive *and* negative fear stimuli; and the fact that their putative high associability affects resistance to extinction, but does not speed up acquisition.

Two more points may be noted. First, the differential associability hypothesis does not fit easily with the neural reality in the amygdala which we believe to lie at the root of both conditioned and innate fear. As noted earlier, what we know of the amygdala suggests that we should see either a quite general increased tendency to fear conditioning for all stimuli, or the presence of a fairly specifically 'wired in' phobic stimulus–avoidance response relationship. Second, 'just as people tend to develop phobias only to certain types of *objects*, they tend to respond to phobia conditioning only when certain types of *fear-evoking stimuli* are coupled with these objects. The observations from the mini-phobia experiments [such as Öhman's] provide the evidence here. Even with snakes and spiders, only stimuli such as electric shocks produce the mini-phobia. Loud noise does *not* produce a fear response to such animals. This result suggests that *only certain sensory pathways* are involved in the acquisition of this type of fear response. Could it be because snakes and spiders bite that tactile stimuli produce the fear?' (Agras 1985, p. 49).

Let us try to produce a unified account able to explain all, or most of, the data. Eysenck's version of Seligman's concept of preparedness is in many ways an amalgam of two separate notions: those of innate reactions to unconditioned stimuli, and of conditioning proper. It is an amalgam, moreover, that is both unnecessary and confusing. The whole point of the concept of conditioning, as Pavlov noted at the outset of his research (Gray 1979b), is to make sense of those cases in which there is *no* biologically prepared connection between the stimulus and the response it elicits. Once there is such a connection, it is hard to see what role conditioning need play. There is plenty of evidence for innate fears of a wide variety of stimuli in a diversity of species (Gray 1987), including fear of snakes in chimpanzees never before exposed to them (Hebb 1946). One of the interesting features of such fears is that they are often subject to maturation; that is, they appear *without learning* at a given stage of ontogeny. This concept is well supported in the animal literature (Gray 1987, Chapter 2) and seems eminently well suited to cope with the evidence that some phobias appear in human beings rather consistently at certain ages (Marks 1969). To account for this age distribution, the preparedness hypothesis must have added to it the ad hoc assumption that conditioning experiences involving the relevant stimuli cluster at the relevant ages. This is no more attractive a notion than the assumption required by the standard conditioning theory that conditioning experiences happen especially often when spiders or snakes are around.

The supposition that the fear-relevant stimuli used by Öhman's group are, quite simply, innate fear stimuli in our species makes immediate sense of several of the findings the group reports (the heightened response to fear-relevant stimuli *before* conditioning occurs; the sensitization by prior aversive stimulation; the greater response at CS onset) and of the characteristics of phobic avoidance (maturation, persistence). It remains to account for the greater discriminability of innate fear stimuli in a Pavlovian conditioning paradigm, and for the greater resistance to extinction of this differential conditioning. To do this, recall that it is the skin conductance response which has been used to demonstrate these effects, and that this has a special affinity for innate fear stimuli. Öhman's group did not obtain similar results using heart rate (Fredrikson and Öhman 1979). As already noted, there are a number of reasons to consider the skin conductance response as an orienting response, unconditioned, sensitized, or conditioned, and probably (in many of Öhman's cases) a mixture of all three. If so, according to the theory developed earlier in this book, this effect of the innate fear stimuli (as opposed to any effects they have on avoidance) would be mediated by the behavioural inhibition system. This suggestion is consistent with the fact that the skin conductance response (unlike phobic avoidance in human beings or active avoidance in animals) is reduced after benzodiazepine or barbiturate administration (Marcy and Quermonne 1974), as well as by septal lesions (Holdstock 1969, 1970). Furthermore, Fowles (1980) has reviewed much of the relevant human literature and concludes that 'electrodermal activity increases when there is activation of the behavioural inhibition system', whereas heart rate does not change.

Thus, we may suppose that innate fear stimuli have two potential effects. First, they act on the amygdala and have the capacity, under suitable conditions, to elicit avoidance responses. Second, they act on the septo-hippocampal system (probably on information relayed to it by the amygdala and as a result of anticipatory conditioning) to generate an orienting response, including the skin conductance component. If we adopt this

hypothesis, we can account for one of the two remaining features of Öhman's results on which the preparedness theory stumbles: the ease with which innate fear stimuli are differentiated into CS+ and CS–. For this is predicted directly from the properties we have given to the behavioural inhibition system. One of the specific functions of this system is to increase level of arousal and spread of attention so as to enhance multidimensional analysis of the stimuli encountered in a threatening environment. Thus, in Öhman's experiments, the stimuli (snakes, spiders, etc.) are both the source and the target of this increased level of processing, and both the cause of increased discriminability and the means through which this discriminability is made evident. The fact that this superior discrimination is allied to greater resistance to extinction of the skin conductance response (as distinct from its speed of acquisition) can also be derived from this approach, if we bear in mind the evidence (Sutherland and Mackintosh 1971; Mackintosh 1974, pp. 438–9) that resistance to extinction increases as a positive function of the number of stimulus dimensions to which the subject attends during training. If it is activation of the behavioural inhibition system which causes an increase in multidimensional stimulus analysis, it follows that there should also be an increase in resistance to extinction of a differential CR.

To sum up, the bulk of clinically reported phobias do not seem to fit with a simple conditioning model of the type suggested by Watson and Rayner (1920; see also Eysenck 1957). The preparedness hypothesis took from conditioning its central position, but still gave it a substantial role in a specialized form. The present position discards the notion of there being specialized forms of conditioning and replaces it with the idea that phobias result from the activation of innate fear responses.

There remains the question why some individuals become phobic and others do not. We have already noted the maturational aspects of phobia; one answer to the question, therefore, is that the factors which give rise to this maturation in a given individual may be more or less extreme for genetic and/or developmental reasons. This fits with the common experience that there are many who are reluctant to approach snakes or spiders or are nervous in the dark, and yet would not be classified as clinically phobic. In the first edition we linked specific phobia with agoraphobia and social phobia, but here we have reason to distinguish them. We do not link sensitivity to phobia, therefore, with the personality traits of neuroticism and introversion, these being linked (Chapter 12) to sensitivity to anxiety. The data on the personality of individuals who have, for example, small animal phobias are consistent with this point of view: unlike people who suffer from other anxiety-related disorders, these individuals are not at the extremes of either neuroticism or introversion (Marks 1969). The variance in phobias attributable to heredity is about 50% (Torgersen 1979), implying that the remaining 50% is attributable to environmental factors.

To understand the heritable component, let us return to the earlier discussion of how an innate fear or anxiety stimulus becomes established in the first place. We suggested that (given a background of genetic variance) there can be an incremental progression from the capacity to form a conditioned response, as a result of the strengthening of connections by experience, to the capacity to demonstrate this response without any prior learning, because of the phylogenetic strengthening of those same connections. 'This fund of structural memory . . . can be properly characterized as *phyletic memory*,

or memory of the species. . . . An unprovable but seemingly reasonable tenet . . . is that the phyletic memory . . . [has] developed in evolution by essentially the same mechanisms of association by co-occurrence and contiguity that contribute to the formation of individual memory in higher cortical systems, although the time scale of the two developments would be immensely different' (Fuster 1995, p. 10; see also Mineka 1992 on 'evolutionary memory'). This leads to two contrasting explanations of existing heritable variation. First, common phobic stimuli might be stimuli which are less than completely established in human populations. Second, and more likely given the primate evidence, common phobic stimuli might be stimuli which used to be more powerful innate fear stimuli in our ancestors, but which have undergone partial 'phyletic extinction' (on analogy with Fuster's phyletic learning) as a result of the increasing defensive capacity of the hominid line. In both cases there will be genetic variation in the population which affects the sensitivity to particular stimuli (and which can be regarded as variation in temperament).

The environmental component of individual variation includes both ontogeny and variations in the conditioning situation. The ontogenetic component involves many processes (e.g. latent inhibition, observational learning, experience of control, experience of 'occasion setters'), all of which affect the future conditionability of specific stimuli, particularly of aversive stimuli (Mineka and Zinbarg 1996, pp. 144–50). Except in very carefully controlled laboratory populations, these processes must vary markedly from individual to individual. We shall see later, when we discuss post-traumatic stress disorder, that ontogenetic events can also affect sensitivity to aversive outcomes in general. The situational component is the mirror of the ontogenetic: the precise stimuli to be conditioned and the context of conditioning affect what is learned. These situational factors will vary between individuals even if their ontogeny did not. In practice, both types of factor will interact to produce individual variations in conditionability.

To conclude that phobia often depends on the activation of innate fear responses is not to say that conditioning is never involved in phobia. As indicated earlier, there are cases in which phobic stimuli cannot in any manner be described as prepared or, by the same token, as reflecting only innate fears (Rachman and Seligman 1976). Eysenck (1977), for example, describes a case in which a man developed a phobic reaction to a pattern of wallpaper which had been on the walls of a bedroom in which he had been set upon by an irate husband. Cases such as these, however, are rare enough and distinct enough to be exceptions that prove the rule. They offer no problem for the theory developed here. As noted above, the persistence of clinical phobic avoidance is what we would expect from the persistence of active avoidance in general. Furthermore, the capacity of simple exposure to eliminate phobic avoidance is what we would expect by analogy with appetitive extinction: loss of response as a result of the omission of the expected reinforcer. It is irrelevant to this subsequent reduction in strength of connections whether the initial strength was due to genetic or environmental factors.

The analysis of avoidance behaviour that has been presented in this book, coupled with the lack of effect of anxiolytic drugs on phobic avoidance (Sartory *et al.* 1990), leads us to expect that, when a person is faced with a phobic stimulus, whether innate or conditioned, the critical sites activated in the brain should include the amygdala and perhaps the hypothalamus or the anterior cingulate cortex. This expectation is supported

by a number of recent *in vivo* neuroimaging studies. The amygdala has been reported to be activated in normal subjects by exposure to fearful facial (Morris *et al*. 1996; Phillips *et al*. 1997) or vocal (Phillips *et al*. 1998) expressions and conditioned fear stimuli (Büchel *et al*. 1998; LaBar *et al*. 1998); and the anterior cingulate gyrus, in normal subjects by exposure to conditioned fear stimuli (Büchel *et al*. 1998) and in phobic subjects during exposure to their feared objects (Rauch *et al*. 1995). These neuroimaging data for the amygdala are consistent with observations of impaired recognition of fearful facial expressions in patients with damage to the amygdala (Adolphs *et al*. 1994; Calder *et al*. 1996).

The role of exposure in extinction may be more easily understood in terms of normal ritual sanitary precautions. Cleansing rituals are clearly worth repeating in the absence of the punisher, disease. Indeed, we attribute the continued absence of disease to the ritual (when we think about it at all). However, if such a ritual were regularly blocked and we found that no disease followed, we would conclude that it was not needed. This type of explanation emphasizes cognitive factors more, and stimulus–response factors less, than in the original learning theory on which behaviour therapy was based. However, as discussed in Chapter 3, the radical behaviourism which gave rise to behaviour therapy has been replaced by cognitive behaviourism (see for example Dickinson 1980). Whether as a result or coincidentally, modern psychological therapy too is often highly cognitive in both theory and practice, and it has been suggested that this cognitive orientation should be taken on board even by those using classical behaviour therapy (Zinbarg 1990). Thus a major, and perhaps often the primary, effect of the various behavioural treatments of phobia is to alter the patient's cognitions, albeit indirectly. This is not a revolutionary proposal, since such cognitive changes form the basis for much conditioning also in many animal species. In the specific phobias, the central cognition relates to the necessity to avoid some relatively discrete source of danger, and does not require the general inhibition of other behaviour that is involved in anxiety as such. Thus, we may see the specific phobias as dependent upon escape and avoidance mechanisms which, according to our theory, are centred on the hypothalamus and amygdala and are not sensitive to anxiolytic drugs. Diagnostic systems should, therefore, categorically distinguish specific phobia from anxiety.

11.14 POST-TRAUMATIC STRESS DISORDER

Post-traumatic stress disorder shares with the specific phobias the feature of stimulus specificity, that is there are clearly defined environmental stimuli to which the disorder relates, but with important differences. The stimulus specificity needs to be considered from two points of view. First, there is the initial trauma that gives rise to the disorder. This is described in the tenth edition of the World Health Organisation (1992, p. 147) international classification of diseases (ICD-10) as an event or situation 'of an exceptionally threatening or catastrophic nature, which is likely to cause pervasive distress in almost anyone (e.g., natural or man made disaster, combat, serious accident, witnessing the violent death of others, or being a victim of torture, terrorism, rape, or other crime.' In contrast, in the case of the specific phobias, as we have seen, it is not usually possible

to identify any such initiating event. Second, the symptoms of post-traumatic stress disorder may subsequently be triggered or exacerbated by stimuli that are associated with, or in some way remind the patient of, that initial trauma (although they may also occur without such an obvious environmental trigger; see the discussion of Adamec's and King's experiments below). It is likely that this second type of dependence upon external stimuli is the result, at least in part, of just those normal conditioning processes which, in the previous section, we were forced by the data to discount in the majority of specific phobias. Thus, post-traumatic stress disorder provides an exemplar of the type of conditioned fear initially seen as the route by which the specific phobias were established.

A further way in which post-traumatic stress disorder both resembles and differs from the specific phobias lies in its symptomatology. Phobias consist essentially in avoidance of the phobic object; if that avoidance is successful, there is no further distress. In post-traumatic stress disorder, there is also strong and persistent avoidance of stimuli associated with the initial trauma, coupled with efforts to avoid all thoughts, feelings, or conversation associated with it. However, there are in addition a range of other symptoms that are not seen in the phobias, and which avoidance of trauma-associated external stimuli is insufficient to control. These include, in particular, recurrent and intrusive distressing recollections of the traumatic event, in the form of thoughts or images; recurrent dreams or nightmares of the event; 'flashbacks', i.e. the experience of reliving the original traumatic episode, often with accompanying illusory or hallucinatory perception; re-enactment of the original behaviour; persistent symptoms of increased arousal and hypervigilance; and a general numbing of emotional reactivity.

In some but not all respects, then, post-traumatic stress disorder shares features with the phobias. There are two features of the condition, however, that bring it closer to panic disorder. The first is the high level of arousal and autonomic disturbance seen in both. The second is the fact that both conditions can manifest themselves during sleep, as nightmares or panic attacks respectively. In the case of relaxation-induced panic, we accepted above Graeff's (1994) hypothesis, according to which this reflects the switching off during relaxation or sleep of descending inhibitory control from the septo-hippocampal system over the periaqueductal grey neurons that act as the final common pathway for panic behaviour. The same hypothesis may also, then, be applicable in some degree to the nightmares characteristic of post-traumatic stress disorder.

This possibility is strengthened by the fact that, like panic disorder (Dantendorfer *et al.* 1996), post-traumatic stress disorder is associated with structural damage in the hippocampus: a decrease in hippocampal volume has been reported in both combat veterans (Bremner *et al.* 1995a; Gurvits *et al.* 1996) and survivors of childhood physical and sexual abuse (Bremner *et al.* 1997; Stein *et al.* 1997). Such damage to the hippocampus may underlie the memory deficits observed in patients with post-traumatic stress disorder. Deficits of this kind have been reported both symptomatically and experimentally. Symptomatically, there can be partial or total amnesia for the traumatic event itself; additionally, gaps in memory have been described in combat veterans as recurring for many years after the Vietnam war (Bremner *et al.* 1995b). Experimentally, impairments have been noted in short-term verbal memory in a group of Vietman veterans with combat-related post-traumatic stress disorder; furthermore, there was a positive correlation ($r = 0.64$) in this group between right hippocampal volume and memory

score (Bremner *et al.* 1995b). This finding stands in need of replication, however, since, in a study of women with reduced hippocampal volume after a history of childhood sexual abuse, no relationship was observed between hippocampal volume and scores on a test of verbal learning (Stein *et al.* 1997). Along with such *deficits* in memory, a key symptomatic feature of post-traumatic stress disorder consists in *intrusive* memories, related to the traumatic incident. This symptom can perhaps be interpreted as reflecting heightened activity in the hippocampal system. There is no reason why such heightened activity should not coexist with partial structural damage in the hippocampus. Indeed, one might see just such a combination as giving rise to a continuous effort towards conflict resolution, entailing reactivation of negatively affective material (perhaps in interaction with the amygdala), but without sufficient processing resources to attain the resolution.

Our discussion of the septo-hippocampal damage seen in panic disorder (Section 11.12) indicated the two most likely aetiologies for this damage as perinatal anoxia or the neurotoxic effects of prolonged exposure to corticosteroids. In the case of post-traumatic stress disorder, the latter is clearly the most likely. The initial trauma is virtually certain to give rise to massive release of adrenocortical hormones, whose levels are likely then to receive a further boost with each re-evocation of the traumatic experience (Bremner *et al.* 1995a, p. 974). The neurotoxic effects of corticosteroids act synergistically with other excitatory inputs to the relevant neuronal targets (namely, pyramidal cells in the hippocampus). Damage to these cells indicates, therefore, that there has been a strong activation of the hippocampus during and/or after the initial traumatic event. The normal function of the hippocampal cellular targets of corticosteroids includes the exercise of a negative feedback over further excessive release of these hormones (McEwen *et al.* 1992). Once these cells are destroyed, further release of corticosteroids is no longer restrained by this feedback. This may explain why patients with post-traumatic stress disorder show abnormally high fluctuations in peripheral levels of cortisol (Southwick *et al.* 1994). Abnormal levels and reactivity of adrenocortical hormones are also likely to affect psychological state directly, increasing in particular the incidence of depression (which, indeed, often accompanies post-traumatic stress disorder). Thus, as in the case of panic disorder, we see again the operation over a long time-scale of a vicious spiral: an initial environmental insult, via prolonged activation of the septo-hippocampal system and stress-induced hormonal changes, can give rise to structural damage in this system, which then exacerbates the physiological and psychological consequences of further exposure to stressful conditions.

This line of argument suggests, then, that the nightmares characteristic of post-traumatic stress disorder, like panic during sleep in panic disorder, may reflect intense activity in a structure at least partially freed from inhibitory control as a result of structural damage to the hippocampus, combined with a further (normal) weakening of this control at the transition from waking to sleep. If this hypothesis is correct, it can perhaps be extended to other symptoms of post-traumatic stress disorder. Consider, for example, flashbacks in the waking state. These phenomena have obvious similarities to nightmares, in that both involve quasi-hallucinatory re-experiencing of events that resemble the original trauma. Thus flashbacks may involve a loss of inhibitory control similar to that which we proposed as underlying nightmares, and with a similar possible relationship

to hippocampal structural damage. This hypothesis is readily testable. It predicts a positive correlation between the extent of reduction in hippocampal volume, on the one hand, and the frequency or intensity of both nightmares and flashbacks, on the other; and, further, that flashbacks (since they do not require the additional weakening of hippocampal inhibitory control that occurs at the transition to sleep) will on average require a greater loss of hippocampal tissue than will nightmares. A further prediction is that any symptoms that reflect in this way loss of inhibitory control should increase with the progress of neurotoxic damage to hippocampal pyramidal cells. Animal studies suggest that this is likely to occur over weeks or months rather than years (McEwen *et al.* 1992). A careful study of the early time course of the appearance of different symptoms after trauma, especially if accompanied by neuroimaging data, would provide a means of testing this prediction.

So far in this account we have concentrated on hippocampal damage, and the possible consequent release of other structures from hippocampal inhibitory control. But which are these other structures? In this context, a putative animal model of post-traumatic stress disorder studied by Adamec (1998a,b,c) is suggestive. This author showed in cats that administration of just one or two systemic injections of an inverse benzodiazepine agonist, FG 7142 (*N*-methyl-β-carboline-3-carboxamide), with known anxiogenic properties in man, is capable of inducing an increase in defensive behaviour that lasts more than 140 days; and demonstrated that this treatment gives rise also to long-term potentiation of responses to electrical stimulation of the amygdala, the augmented responses being observed in both the ventromedial hypothalamus and the periaqueductal grey. Adamec advances a number of reasons to believe that the latter, electrophysiological changes underlie the increased behavioural defensiveness caused by FG 7142. In a somewhat similar study, King (1999) showed that electrical stimulation of the rat's superior colliculus (an important component of the visual system) leads to a long-lasting increase in defensive behaviour that she too likens to post-traumatic stress disorder. The observed changes included intense visual scanning and preparations to engage escape behaviour. The visual scanning is consistent with the known role of the superior colliculus in visual exploration. In addition, the projection from the superior colliculus to the periaqueductal grey (King *et al.* 1996) is likely to be of importance in the overall pattern of heightened defensive behaviour. Both Adamec's and King's experiments suggest that the emphasis placed above on the conditioning origins of at least part of the symptoms of post-traumatic stress disorder probably needs to be supplemented by an additional process, namely, a 'lasting transformation in emotional disposition' (Pitman *et al.* 1993), that is, a general increase in the propensity to react with excessive defensive behaviour to all kinds of threat whether or not they closely resemble the original traumatic event.

The equivalents of enhanced defensiveness in the symptomatology of post-traumatic stress disorder probably include (as listed in DSM-IV; American Psychiatric Association 1994) avoidance of stimuli associated with the trauma, exaggerated startle response (demonstrated also in laboratory experiments in the relevant patient group by Butler *et al.* 1990), irritability or outbursts of anger and hypervigilance. In line with the general picture of primary defence systems presented in this book, we may attribute the avoidance behaviour to the connections between the amygdala and the medial hypothalamus. The remaining symptoms in the DSM-IV list can plausibly be related to the output

from the amygdala (and perhaps, in the light of King's observations and especially in the case of visual threat, from the superior colliculus) to the periaqueductal grey. A further symptom that can perhaps be attributed to activity in the latter region is that of emotional numbing. It has been suggested by van der Kolk *et al.* (1989) that this may be analogous to, and perhaps even identical with, analgesia resulting from stress-induced release of endogenous opioid transmitters, a phenomenon well established in the animal literature (Akil *et al.* 1984). In support of this hypothesis, van der Kolk *et al.* (1989) report that combat veterans with post-traumatic stress disorder, while viewing combat-related but not neutral videos, had elevated pain thresholds (to heat stimuli) compared to control veterans without this diagnosis; and that this analgesic response was abolished by systemic administration of the opiate receptor antagonist, naloxone. In further support, Baker *et al.* (1997) measured β-endorphin in cerebrospinal fluid and found elevated levels in patients with post-traumatic stress disorder as compared to normal controls. There is evidence that this type of opioid stress-induced analgesia is mediated in the periaqueductal grey in both animals (Fanselow 1991) and man (Hosobuchi 1981); and both this region and the amygdala are rich in opiate receptors. Thus, these putative roles for the periaqueductal grey in mediating the symptomatology of post-traumatic stress disorder provide a further parallel to the analysis of panic described in Section 11.12.

However, the periaqueductal grey, presumed critical site for panic attacks, does not look plausible as a location for producing all the symptoms of post-traumatic stress disorder. In particular, flashbacks and nightmares are strongly perceptual in nature, suggesting a forebrain location. As we saw in Chapter 6, as well as its connections downstream to the hypothalamus and central grey, the amygdala has strong connections also with sensory systems. A reasonable hypothesis, therefore, places the perceptual phenomena of flashbacks and nightmares in these connections. Hippocampal inhibitory control (or, in post-traumatic stress disorder, the loss of such control) could then influence the intensity of these perceptual phenomena either by way of the connections from the hippocampus to the amygdala itself, or by way of converging connections from the amygdala and the hippocampus to neocortical sensory-processing areas. This hypothesis—that the amygdala plays a critical role in the laying down and retrieval of strongly emotional memories—has been proposed by both LeDoux *et al.* (1989) and Pitman (1989). It is supported by the results of recent human neuroimaging studies. Using positron emission tomography, both Rauch *et al.* (1996) and Shin *et al.* (1997) report increased relative regional blood flow in the amygdala when combat veterans with post-traumatic stress disorder engaged in trauma-related imagery. In the same studies, increased blood flow was also seen in the anterior cingulate gyrus, consistent with the role allotted to this region in our general treatment of the primary defence system. Note again the similarities between these results and those obtained in similar studies of simple fear conditioning and the specific phobias (see Section 11.13.2). A further interesting observation in the Rauch *et al.* (1996) study was activation of secondary visual cortex, consistent with the predominantly visual modality in which trauma-related imagery was evoked and with the proposal that flashbacks are mediated by amygdalar connections to sensory neocortex.

An alternative, but not mutually exclusive, way of accounting for flashbacks that puts together many of the same neuropsychological ingredients has been proposed by Nadel

and Jacobs (1996). These authors point to evidence that high levels of stress-induced corticosteroids reduce hippocampal excitability while at the same time potentiating amygdalar function. These phenomena, they suggest, could impair, at the time of initial exposure to trauma, the encoding (hippocampally mediated) and subsequent storage of information about context, while leaving intact encoding (via the amygdala) of associations between specific cues and the traumatic event. In consequence, 'when a person retrieves a traumatic event memory, the retrieved information is bereft of . . . context . . . [and] the memory takes on a quality of the here and now so strongly that the individual may literally *re-experience the event*' [their emphasis]. Subsequent steroid-induced structural damage to the hippocampus might exacerbate the effects of these alterations in encoding processes by further weakening the capacity, at the time of retrieval, to analyse contextual differences between the past and current environments.

11.15 AGORAPHOBIA

Whether or not conditioning is necessary for the production of simple phobic reactions, it undoubtedly plays a role in extending their range. Thus, someone who is afraid of spiders may curtail activities which bring them into contact with stimuli only secondarily associated with spiders. Similarly, one can acquire a conditioned fear of a discrete stimulus such as a gun, or generalize a fear of snakes to, for example, a crocodile skin handbag. Associations of this kind may spread along the entire intricate network of routes offered by stimulus and semantic (Razran 1971) generalization. However, this does not mean that we should jump to the conclusion, as in the first edition, that agoraphobia and social phobia are essentially simple phobias of particular types of stimuli. This change in perspective is driven by the material dealt with in Chapters 2 and 4 of the present edition. We deal, then, with these two types of 'phobia' next, arguing that they have been misnamed and should in fact be classified as types of anxiety.

A particularly important way in which the conditioning process comes into play lies in the development of anxiety in relation to situations associated with the physiological consequences of fear itself. We say 'anxiety' here, because in most cases we are dealing with cases of passive avoidance of potential, often unpredictable, negative events rather than active avoidance of discrete, localizable, actual negative events. In many cases, this appears to be the way in which agoraphobia is generated. For example, a person who experiences a panic attack in a particular environment may develop conditioned anxiety to, and avoidance of, that environment. The fact that the panic often cannot be attributed by the patient to a discrete, localizable external stimulus increases the chances of associational spread of the conditioned response, with the resultant development of a generalized agoraphobia. Open spaces may act like this, as a diffuse set of poorly localized environmental stimuli; or, alternatively, they may themselves represent a source of uncertain threat able to act as an innate anxiety stimulus requiring risk assessment and behavioural inhibition. In either case, such spatially diffuse stimuli are particularly likely to activate the septo-hippocampal system, as initially proposed (as we have seen, with much subsequent empirical support) by O'Keefe and Nadel (1978). Crowded places similarly provide a complex set of diffuse and poorly localized social stimuli as a source

of indefinite potential threat. The key point, given an analysis in terms of defensive direction, is that the specific phobias in general manifest as increased active avoidance of the specific phobic object while agoraphobia manifests as increased passive avoidance, inhibiting entry into appetitive situations because of a concurrent associated threat.

Given our theory, panic attacks in the classical sense would not be the only possible UCS at the origin of conditioned agoraphobia. Any other spontaneous, unpredictable aversive reaction could in principle produce the same conditioning. 'Agoraphobia without a history of panic disorder' seems in many cases to be the result of 'limited symptom' attacks which do not fulfil the criteria for full-blown panic, but may well provide a basis for conditioned avoidance (Goisman *et al.* 1995). Well-established avoidance could then reduce the occurrence of the feared internal response through a general reduction in the level of arousal (see below).

As outlined in detail in Section 11.6, agoraphobia with panic attacks can arise either from a primary panic disorder, with resulting conditioning to the environment in which the panic occurs, or from a primary heightened anxiety that leads to the occurrence of panic attacks. This flexibility in the theory does not necessarily imply lack of testability. It should be possible, for example, to distinguish the types via carefully designed challenge tests. Manipulation of the partial pressure of CO_2, for example, could well determine whether an agoraphobic is particularly sensitive to panic attacks. Similarly, administration of cognitive tests could determine whether the patient is particularly sensitive to anxiety-provoking stimuli. These kinds of test are discussed further in Section 11.20.

11.16 SOCIAL PHOBIA

The emphasis in the first edition was on conditioned stimuli as the major trigger for anxiety. Nonetheless, we recognized that one of the adequate inputs to the behavioural inhibition system was what we then termed 'innate fear stimuli' (and see Gray 1987b). Given the careful analysis provided by the Blanchards (Chapter 2), we now recognize that there can be both innate fear stimuli (e.g. a cat for a rat) and innate anxiety stimuli (e.g. the smell of the cat), and the pharmacological evidence shows that it is only the latter which provide input to the behavioural inhibition system. In discussing social phobia, therefore, the first edition noted an innate component, equated this with the innate fear stimuli of specific phobia, and remained happy with the term phobia as applied to 'social phobia'. Here, however, we reverse this categorization.

Social phobia

> is frequently multiphobic: fears of specific scrutiny situations (e.g. eating, writing, public speaking, appearing in public places, using public toilets) often co-existing together, and there may be a more pervasive anxiety about social interactions present. . . . A recent report indicates that two thirds of a large clinical sample suffer this generalised form of the disorder, in addition to fears of multiple specific situations. . . . While it seems likely that social phobia can occur in a 'pure' form (e.g. public speaking anxiety), for many sufferers [social phobia] and avoidant personality disorder are probably disorders on the same continuum. . . . In terms of other problems associated with social phobia, research has indicated that self-medication using alcohol, or other anxiolytics is frequent, and it is not unusual for this self-medication to escalate into significant substance

abuse. . . . The presence of significant depressive symptoms is reported to be quite common, being found in half the patients in one study . . . There are frequently clinically significant levels of generalized anxiety disorder. (Mattick 1990, pp. 180–1.)

The lack of specificity of stimuli, the fact that where specific stimuli are involved we are dealing with anxiolytic-sensitive reactions, the fact that anxiolytics are used as self-medication, and the presence of depression as a, presumably long-term, consequence all fit with the view that social phobia would be better termed social anxiety. What, then, is its source?

Stimuli that arise in the course of social interaction are among the most important innate threats in many animal species (Gray 1971); and social phobia (anxiety) appears to be a response to such stimuli (Marks 1969). Furthermore, diagnosis of social phobia as a disorder appears to depend on a quite arbitrary threshold to select an extreme portion of a continuous distribution which includes the normal population (Stein *et al.* 1994b). In one of Öhman's experiments using non-clinical subjects, for example, pictures of angry faces were classified as fear-relevant and pictures of faces with neutral or happy expressions as fear-irrelevant by the same criteria as pictures of snakes compared to pictures of flowers (Öhman and Dimberg 1978). It was proposed by Gray (1976) that 'stimuli arising in the course of dominance interactions' are among the kinds of stimulus which activate the behavioural inhibition system. It is relevant, also, that septal lesions and anxiolytic drugs increase social interaction (Appendices 1 and 8).

However, while anxiety is, according to our theory, dependent upon activity in the septo-hippocampal system, this is not where we would locate the primary pathology in social anxiety disorder. Rather, we expect this to lie in some part of the primary defence system (such as the amygdala or anterior cingulate cortex), giving rise to either conditioned or innate overreaction to socially threatening stimuli. Consistent with this view is the evidence, noted above in the section on simple phobias, that the amygdala plays a central role in the perception of both facial (Adolphs *et al.* 1994; Calder *et al.* 1996; Morris *et al.* 1996; Phillips *et al.* 1997) and auditory (Phillips *et al.* 1998) expressions of fear. But, if these stimuli are received by the brain in the same way as the innate fear stimuli of simple animal phobias, why are we dealing with anxiety rather than phobia? The answer to this question lies in a fundamental tenet of the theory espoused in this second edition: fear or phobia is a reaction to a set of stimuli that require exit from a dangerous situation; anxiety is a reaction to a set of stimuli that require entry into the situation, with resulting conflict. Specific phobias are correctly so-named because the predominant required reaction is avoidance of a localized source of definite danger. The far-from-specific (Mattick 1990) social phobias are more correctly termed 'social anxieties' because the predominant required reaction is entry into a diffuse source of potential danger admixed with a number of the most powerful rewards available to our species. Thus we attribute to the social anxieties the same innate or acquired strengthening of connections (in some as yet undetermined locations in brain, probably the amygdala or cingulate cortex) as is postulated to occur in the amygdala in the case of specific phobias. However, in the case of the specific phobias, involvement of the septo-hippocampal system (via input from the amygdala) is limited to a modest orienting response. In the case of the social anxieties, involvement of this system is much more

extensive because of the concurrent appetitive qualities of the aversive stimuli. This results in inhibition of the prepotent approach responses (i.e. inhibition of social involvement), as well as increased arousal and attention to negative stimuli (in this case negative social stimuli).

As would be expected from this analysis, the role of environmental factors in determining sensitivity to social phobia seems to be generally similar to the role of those controlling specific phobias (Mineka and Zinbarg 1995, 1996, pp. 157–66). However, 'social anxiety should covary significantly with nonsocial forms of anxiety, and behavioural inhibition may be the temperamental vulnerability factor common to most if not all of the anxiety disorders, including social phobia' (Mineka and Zinbarg 1996, p. 162). Thus, the role of conditioning is relatively non-specific with respect to the nature of the elicited reaction, whereas the type of innate threat stimulus is relatively specific in this respect.

Social anxiety can interact with panic in a manner similar to that seen in agoraphobia (see above), although the panic symptoms reported by social phobics differ from those of agoraphobics (Page 1994), having the appearance more of consequences of high levels of social anxiety than causes. However, another possibility is that panic disorder may be primary in both conditions. Most likely (as we have suggested for agoraphobia) is the possibility that social phobia, as normally diagnosed, may include two or more quite different disorders. In this context, it is relevant that, while social phobia can be ameliorated by anxiolytic drugs, it appears to show a particular sensitivity to selective serotonin reuptake inhibitors (Van Ameringen *et al.* 1994). Thus, while excess anxiety appears to be characteristic of the disorder, the anxiety may in some cases be secondary to disturbance (possibly panic disorder) in some other part of the defence network.

The theory presented here, then, suggests an important similarity between social phobia and agoraphobia: they are both anxieties rather than phobias; and both may depend on an excessive reaction to innate anxiety stimuli. The theory also suggests an important difference between them. Agoraphobia often appears to be an incidental consequence of the coincidence of a randomly occurring panic attack with the presence of otherwise affectively positive stimuli. Social phobia, in contrast, involves social situations in which, as in the case of novel stimuli, both positive and negative affective values can be provided by one and the same object.

11.17 GENERALIZED ANXIETY DISORDER

'Generalized anxiety disorder was originally defined in DSM-III as a residual category. The hierarchical exclusion rules used in DSM-III obscured the independence of generalized anxiety disorder and were dropped in DSM-IIIR, thus allowing an independent diagnosis of generalized anxiety disorder in addition to other mental disorders and recognizing the importance and relative frequency of psychiatric comorbidity' (Brawman-Mintzer *et al.* 1993, p. 1216; see also Brown *et al.* 1994 for a consideration of the weaknesses of definition of this condition). While there is very high comorbidity, it also seems clear that generalized anxiety can occur in isolation, with the same symptoms whether it is isolated or comorbid (Brawman-Mintzer *et al.* 1993, 1994; Wittchen

et al. 1994). For our purposes, this is fortunate since, given the analysis so far, generalized anxiety disorder is, in essence, the only clearly identifiable primary anxiety disorder. That is, it is a case of maladaptive anxiety in which the primary pathology lies in the control of anxiety itself, as opposed to some mechanism (panic, phobia, innate stimulus detection) which then has consequential effects on the central control of anxiety.

Under normal circumstances, anxiety requires some eliciting source of conflict. According to our theory, the septo-hippocampal system will receive information about concurrently activated, conflicting, goals and this will result in inhibition of prepotent behavior and, especially in cases involving aversive motivation, increases in arousal mediated by the amygdala. In all of the cases we have considered so far, there has been an abnormality or oversensitivity of some area other than the septo-hippocampal system which has increased or replaced a normal aversive input and so engendered the resultant anxiety. By contrast, if there is septo-hippocampal overactivity (or overactivity in those portions of the amygdala recursively connected with the septo-hippocampal system; see below), we would expect suppression of prepotent behaviour and the substitution of risk assessment, coupled with an increase in the relative weighting of affectively negative associations.

Increase in affectively negative cognitive bias is particularly important in this context. Approach–avoidance conflict could be resolved by suppression of either approach or avoidance. However, it is fundamental to our theory (based as it is on the actions of the anti-anxiety drugs and on the role of interference in hippocampal amnesia) that the business of the hippocampus is to increase the power of negative at the expense of positive associations, and so to produce avoidance rather than approach. Conversely, it is by decreasing activity in the septo-hippocampal system that anti-anxiety drugs decrease avoidance. The putative overactivity of the septo-hippocampal system in patients with generalized anxiety disorder is consistent with the frequent inability of such patients to identify an explicit source of threat (in contrast, for example, to the panic attacks that can often be identified in agoraphobia). Such overactivity is consistent also with the general picture of cognition in these patients. 'It has long been known that anxiety patients are pre-occupied with thoughts about threats, and that they have unrealistic beliefs about danger. There is also substantial evidence that correcting these distortions can decrease anxiety. Also, anxiety patients preferentially attend to and remember threatening cues. They interpret ambiguous words as threatening. They are slower to name colors of threatening words in the Stroop task, and show increased interference to threat words in a subliminal dichotic listening task. These biases are not entirely state dependent, since diazepam reduces anxiety in outpatients with generalized anxiety disorder but does not change their bias to attend to threatening material on a Stroop color test' (Nesse and Klaas 1994, p. 465). Non-anxious individuals also 'have systematically distorted perceptions of risk . . . [but, in this, they and anxiety patients are] essentially identical. In confirmation of previously reported patterns of risk distortion, both groups overestimated rare risks, underestimated common risks, and consistently underestimated the relative risks to self. These results suggest that cognitive assessments of [the probability of] specific risks are normal in patients with anxiety disorders' (Nesse and Klaas 1994, p. 465). Thus, it is the assessment of the extent of danger or aversiveness of individual stimuli, or their relative salience, that is abnormal in anxiety patients.

Why should overactivity in the septo-hippocampal system necessarily result in aversive output when we have argued that many of its normal operations involve the resolution of conflicts that are not particularly threatening? There are several possible answers to this question. The simplest is that, if activity in the septo-hippocampal system is *generally* increased, this will lead to excess output which will be unusual in going to all its targets, including the amygdala. The resultant increased activity in the amygdala (the source of the arousal output of the behavioural inhibition system) could then supply the affectively negative component of the condition. This is the alternative which, at present, we favour. A second possibility is that generalized anxiety disorder involves increased activity specifically in that part of the hippocampus which projects to the amygdala, or in that part of the amygdala which projects to the hippocampus, or in the interconnections between the two. In sum, generalized anxiety disorder reflects overactivity in one or more of the parts of the CNS devoted to the resolution of conflict, behavioural inhibition, and risk assessment. At the present stage of development of the theory, it is most parsimonious to assume that this disorder is specific to overactivity of the septo-hippocampal system. Given the cognitive functions that we have ascribed to this system, this position is very close to the views of Mathews and of Michael Eysenck discussed above.

11.18 OBSESSIVE–COMPULSIVE DISORDER

Having discussed one set of cognitive biases which seem fundamental to generalized anxiety, we turn now to another set which we believe has been confused with it: that of obsessive–compulsive disorder. In the clinic, obsessive–compulsive disorder is usually accompanied by anxiety. Conversely, normal anxiety (as of the mother for the safety of a child in a dangerous place) frequently gives rise to repetitive checking behaviour. After all, checking is one of the primary functions of the behavioural inhibition system. However, obsessions and compulsions can occur in the absence of anxiety and so, as with panic, need to be kept logically separate from it (this is an important departure from the first edition).

To most of us the symptoms of obsessive–compulsive disorder are more mysterious than those of the specific phobias. If you say you are frightened of snakes, you are likely to be viewed as timid but rational. However, if you say you must wash your hands exactly 84 times and must do this every hour, you risk being viewed as quite mad. Conditioning theories of the origin of anxiety treat the compulsions as active avoidance behaviour (and so do we, but in a somewhat different fashion), maintained by a reduction in anxiety that follows them. As expected on the active avoidance model, environmental trigger stimuli (e.g. dirt in the case of hand-washing) provoke the ritual in many cases, exposure to such stimuli without performance of the ritual increases subjective anxiety, and performance of the ritual is followed by a decrease in anxiety (Hodgson and Rachman 1972; Röper and Rachman 1976). The relationship with phobia, which we have also interpreted as a form of active avoidance behaviour, extends to the resistance of compulsions to anxiolytic treatment and the fact that systematic exposure to 'anxiogenic thoughts' (equivalent, given the active avoidance model, to the phobic stimulus) tends to alleviate the compulsive rituals (e.g. Lovell *et al.* 1994).

However, obsessions call for a different type of explanation, as does the fact that (as with simple phobic stimuli) the rituals seen in this syndrome are not a random sample of the kinds of behaviour which could just happen to precede a reduction in anxiety; nor are obsessions a random sample of the ideas which could be threatening. There is great similarity in the kinds of ritual and obsession observed in widely differing places. In New Delhi (Akhtar *et al.* 1975) as in England (Rachman 1978), the commonest obsession is with dirt, disease, or contamination (50–60% of reported obsessions), followed by orderliness or aggression (20–35%), and then by religion and sex (5–15%). Behavioural rituals similarly consist, in the great majority of cases, of cleaning, tidying up, and checking that various potential threats are absent.

Let us leave the notion of active avoidance *conditioning* as the original source of obsession and compulsions, and consider instead the related notion that the behaviour is active avoidance *behaviour*, but not resulting from conditioning. Let us suppose that, as with specific phobia or social anxiety, we are dealing with innate defensive stimuli (see also Jakes 1996, especially p. xvii). What kind of stimuli might these be?

Dirt is one of the commonest obsessional preoccupations (Akhtar *et al.* 1975; Rachman 1978), and it is likely that natural selection would reinforce the occurrence of grooming behaviour in response to dirt. Indeed, in the case of organisms able to cause disease, grooming and cleanliness, especially washing, in the total absence of any perceptual evidence of contamination, would also be adaptive. This should favour, then, the spontaneous appearance of the same cleaning rituals as would be desirable in response to detectable contamination. In at least some cases, therefore, it is likely that obsessive–compulsive disorder involves the excessive activation of normally adaptive innate avoidance behaviour (and its accompanying cognitive complements) for which there can, under normal ecological circumstances, be no external safety signal to terminate the behaviour. The lack or weakness of such external signals offers a natural account of the fact that hand-washing, checking that the door is locked, and so on are hardly ever sufficient to reduce anxiety in these patients if they are done once; indeed, it is commonly found that they must be done tens or even hundreds of times.

If obsessions and compulsions reflect innate avoidance tendencies, this does not link them directly with anxiety or even fear. As discussed earlier, normal avoidance behaviour, once it is established and successful, is produced in response to the CS in the absence of any obvious fear and can be viewed as a simple habit (for a useful critique of the different view we took in the first edition, see Jakes 1996, pp. 52–6). This should be as true of phylogenetically strengthened tendencies as associatively strengthened ones. In some cases, therefore, especially early in the progression of the syndrome, we would expect ritual behaviour to occur in the absence of anxiety or fear (except that generated by the senseless intrusions and uncontrollable obsessions themselves). However, blocking of the ritual (as in blocking of a well-learned avoidance response) would be expected to result in fear. Furthermore, as occurrences of the ritual increase in frequency, its interaction with normal daily activity would have the consequence that it comes either to be blocked or to produce interference with the normal activity, both of which could be equally disturbing. At this point we would expect the sufferer to seek clinical help.

Does this mean, then, that anxiety is simply a symptom of a more fundamental obsessive–compulsive disorder? No more so than that anxiety is simply a symptom of panic

disorder. One of the main functions of the behavioural inhibition system is to produce and control checking behaviour. High levels of anxiety would, then, be expected to produce some obsessional ideation and compulsive ritual behaviour. Thus, there should often be positive feedback between obsessions and anxiety, analogous to the relationship between panic and anxiety. Pursuing the analogy further, there may be symptomatically similar patients who are primarily obsessional or primarily anxious, respectively; but it seems likely that in many cases obsessive–compulsive disorder will be seen clinically in those who have a heightened sensitivity to both anxiety and obsession.

Let us see how this analysis maps onto the available neurology. As concluded in Chapter 6, following Rapoport (1989), we can identify obsessions with activity in the cingulate cortex (or the prefrontal cortex; e.g. Swoboda and Jenike 1995), and compulsions with excessive output from the basal ganglia (which, given the essentially inhibitory character of basal ganglia outputs, may imply *decreased* neural activity; Lucey *et al.* 1997a,b). To some extent a system structured in this way can be viewed as autonomous in the normal production of a variety of innate motor programs. However, we have included the posterior cingulate in the septo-hippocampal system on the basis of its anatomical connections and its theta activity, and we attribute to this region at least part of the active risk assessment which can result from output from the hippocampal formation. We would expect from this neurology, as suggested above, that anxiety can be a source of obsessional checking. But checking as such would remain located in the cingulate, not in the septo-hippocampal system. This is consistent with the fact that anxiolytic drugs are not generally effective for the treatment of obsessive–compulsive disorder *per se* (Appendix 1), although they would be expected to reduce any secondary anxiety consequent on this disorder (see below).

For both functional and neurological reasons, then, we arrive at the following view of obsessive–compulsive disorder. Obsessions and compulsions can arise from overactivity in the cingulate–basal ganglia circuitry. Often this will simply give rise to avoidance behaviour (or successful checking), will not produce major increases in anxiety, and will not lead to the seeking of clinical help. If rather less coherent components of the motor programs designed to cope with threatening stimuli are activated, this might be expected, as described by Rapoport (1989), to give rise to restricted and relatively anxiety-free cognitive variants of the motor tics characteristic of basal ganglia structural disease. Given current diagnostic categories, such conditions would still be classified as obsessive–compulsive disorder. At the other end of the spectrum, where the frequency of fully fledged avoidance behaviour is very high, or the adequate stimuli elicit avoidance in problematic situations, or the avoidance response is in some other way blocked, then there will be conflict between this response and others. This conflict is likely to activate the septo-hippocampal system, with consequent anxiety. Such anxiety should be most likely (or will occur at a lower threshold) in those with a neurotic introvert personality. This is one scenario, given the normal progression of symptoms. However, the cingulate is a direct target of the septo-hippocampal system. This suggests an alternative scenario: pathologically maintained anxiety (most likely generalized anxiety) could provide an unusually high level of input to the cingulate-basal ganglia circuitry (via the subicular input to posterior cingulate). This input could then trigger any latent obsessions and/or compulsions.

11.19 A TYPOLOGY OF THE ANXIETY-RELATED DISORDERS

This chapter has provided both ethological and neural (Fig. 11.1) reasons for distinguishing a variety of specific anxiety-related symptoms and syndromes. However, given the distinction between symptoms and syndromes, the use of our clinical typology is not quite so straightforward as the functional typology of defence (Fig. 11.2) on which it is based.

Panic disorder is viewed as the inappropriate production of a reaction normally generated by proximity to an unavoidable major threat or by a stimulus, such as suffocation, which normally produces undirected escape responses. In spontaneous panic, and in some situations in the wild, freezing can occur rather than active escape attempts. The primary control centre for panic as a symptom and the primary location of dysfunction in 'pure' panic disorder is the periaqueductal grey. The intensity of the syndrome may in some cases or at some times reflect loss of hippocampal inhibitory control of this structure.

Specific phobia is viewed as a reaction normally generated by proximity to an avoidable major threat, or a stimulus which predicts its presence, and which normally produces directed escape or active avoidance. The primary control centres for this are likely to lie in the medial hypothalamus and amygdala. The commonest cause of specific phobia, as a disorder, is likely to be a greater than average strength in the connections between innate threat stimuli and their associated set of innate response production systems. Specific phobia can also result (although more rarely) from conditioned avoidance. The occurrence of the phobic avoidance response prevents its extinction; exposure to an innate or conditioned stimulus which predicts, but is not in fact followed by, threat facilitates extinction.

Post-traumatic stress disorder is most simply viewed as resulting from the extreme situation facing the individual. The symptoms that ensue appear to reflect a long-lasting conditioned response involving intense activity in all regions of the primary defence system, namely the periaqueductal grey, hypothalamus, amygdala, and anterior cingulate; combined in some cases with loss of hippocampal inhibitory control, due to structural damage arising most probably from excessive and prolonged action of adrenocortical hormones within the hippocampus. Even given the traumatic circumstances that lead to this condition, differences in pre-trauma personality play a role (McFarlane 1989), as we discuss further in Chapter 12; conversely, the pervasive nature of the changes that ensue, including the structural damage sometimes affecting the hippocampus, represents in some degree a long-lasting change in personality.

Agoraphobia ('agora-anxiety') is viewed as a reaction most frequently resulting from the conditioning of anxiety by spontaneous panic, but which may also result from excessive responsiveness to a diffuse set of poorly localizable, innately aversive spatial or social cues. Here the primary site of excess activity is likely to be in the septo-hippocampal system and amygdala; and the distinctive behavioural pathology, largely the result of conditioning and secondary to the panic attacks or other phobic stimuli which provided the reinforcement for that conditioning.

Social phobia (or better, social anxiety) is viewed as a reaction normally generated by proximity to, or placement in social interaction with, conspecifics. This involves conflict

between alternatives. Such conflict may be between alternative goal objects or between alternative approach and avoidance tendencies to a particular goal object. The primary control centres (i.e. the location of the mechanisms which detect and analyse social situations) are not clear, but could well include the amygdala and cingulate or prefrontal cortices. However, the primary site of pathology could be located either in these centres or in excessively strong connections from there to the septo-hippocampal system. In either case there is likely to be excessive activity in the septo-hippocampal system.

Generalized anxiety is the one case which we see not only as being 'pure' anxiety disorder but also in which the anxiety itself is pathological as opposed to being a reaction that would be normal if its immediate trigger (panic/fear) were normal. There is no relevant ethological analysis here. However, neurological considerations suggest that there could be two different types: anxiolytic-sensitive (presumed to reflect overactivity in the septo-hippocampal system and ameliorated by the action of anxiolytics on ascending modulatory inputs); and anxiolytic-insensitive (depending on descending inputs to the septo-hippocampal system and/or the amygdala from the prefrontal and/or cingulate cortex; see Chapter 6).

Obsessions and compulsions, lastly, can be viewed as innate rituals generated both by specific environmental stimuli and, particularly, spontaneously. In the latter case they normally fulfil the role of defending the animal against threats which cannot be detected, such as disease organisms. We link these to the various checking functions of the septo-hippocampal system, particularly the risk analysis output. However, the key control centres are the cingulate (obsessions) and the basal ganglia (compulsions), and it is here that any neuropathology would occur in 'pure' cases (lacking anxiety), which would not normally present in the psychiatric clinic. Neurally induced spread of activity or conditioning can result in increased anxiety (reflecting increased output from the septo-hippocampal system).

How is it that this inherently simple classification is not reflected in clear and discrete entities in the clinic? There are two reasons, one situational and one neurological.

In terms of situations, while the Blanchards' analysis allows one to categorize reactions in terms of defensive distance, it gives no guarantee that specific systems will be activated one by one. There is good evidence for blending of concurrent emotions (McNaughton 1989, Chapter 4); and, in an example we have used before, a rabbit faced with a fox in its vicinity must be prepared both to remain still and assess whether it has been detected (with the fox as a potential predator) and to flee (with the fox as an actual predator). Thus both anxiety and fear will be present (fear here implying a high tendency to engage active avoidance).

In terms of neurology, while we can equate individual symptoms with activity in individual areas, this gives no guarantee that the site producing the most evident symptoms is the critical site from a pathological point of view. There are delayed links through the environment which are important for aetiology (e.g. the analysis above of the conditioning of agoraphobia by spontaneous panic); and there are neural links within the brain (so that, for example, high levels of anxiety can trigger panic). In some cases, inhibitory connections may blur the issue. In Graeff's view, the presence of moderate anxiety is generally inhibitory of panic, with resultant spontaneous panic attacks during sleep or relaxation. In this case, then, a secondary agoraphobia may partially suppress

(via the serotonin system) the appearance of the panic attacks which were primarily responsible for the conditioning of anxiety in the first place.

We must, therefore, as does modern cognitive psychology, take a more recursive, network view of anxiety and its related phenomena. Under normal circumstances, in an anxiety-provoking situation, there will be some activation of each of the periaqueductal grey, hypothalamus, amygdala, septo-hippocampal system, cingulate–basal ganglia loop, and prefrontal cortex. The precise balance of activity between these areas will be determined by their interconnections and by the modulating influences of acetylcholine, noradrenaline, and serotonin, as well as a constellation of pituitary–adrenal hormones. Thus, in the majority of cases, a primary disturbance in any one area, whether of endogenous or environmental origin, will tend to spread to others, giving rise to a mixture of symptoms (Fig. 11.3). It follows that the primary disorder may not be obvious simply from the symptoms exhibited (even ignoring the effects of culturally specific display rules); and that even if our typology of anxiety disorders were entirely correct, it would still not be easy to determine the type applicable to a particular patient. We deal with this issue in the next section.

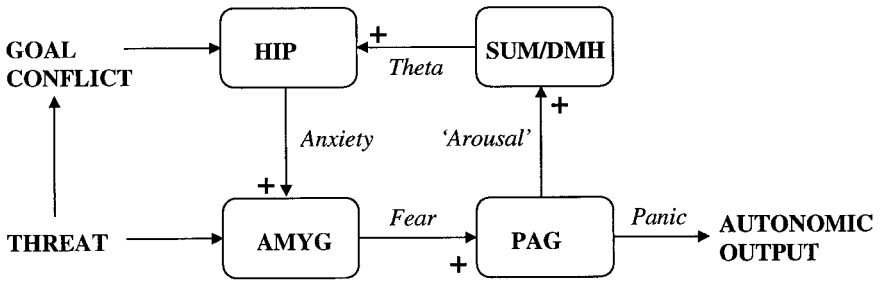


Fig. 11.3 One example of possible positive feedback loops involved in the control of anxiety. HIP, hippocampal formation; SUM/DMH, medial supramammillary and dorsomedial hypothalamic areas involved in the control of theta frequency; PAG, periaqueductal grey; AMYG, amygdala.

11.20 DIFFERENTIAL DIAGNOSIS

Our analysis of clinical anxiety has both optimistic and pessimistic consequences for diagnosis, and hence eventually for treatment. On the optimistic side, the separation of symptoms from syndromes, coupled with the conclusion that syndromes cannot be distinguished simply in terms of symptoms, allows for a resolution of a number of controversies in the literature. Usually both sides are right: panic can be both the cause and the result of anxiety; anxiety can be both the source and the consequence of obsession. On the pessimistic side, we have largely explained away, or explained post hoc, rather than truly explained the clinical situation. This is not to say, however, that the theory is compatible with *any* data that might be collected in future. An important conclusion from our analysis is that it should be possible to separate the different syndromes by

using theoretically based challenge tests and ignoring the fact that different syndromes can present with much the same symptoms. Indeed, a key feature of the tests we propose is that they should seldom be directed towards the most obvious symptoms and should be administered when state anxiety and hence symptoms are minimal. The central idea behind our suggestions for differential diagnosis is that the specific nodes of the defence system should be selectively challenged to determine whether they are functioning normally, and that such challenges should be designed to produce *minimal* reactions from the rest of the defence system. Otherwise, given the recursive interactions characteristic of the defence system, simply making someone anxious (or fearful or panicky) will automatically spill over into activation of much of the system, so making it impossible to determine at which point excessive reactions begin. An important corollary of this recursiveness (and an idea gradually creeping into conventional diagnosis) is that there is likely in any case to be extensive genuine comorbidity. For there is little reason to suppose that just one node of the overall defence system should often be the only one over-reactive in any one individual at any one time. Indeed, given what we saw in the section on post-traumatic stress disorder, it is surprising that over-reactivity is not usually more diffuse and widespread.

Let us now look at some possible challenge tests and ways in which they might be put together to form a diagnostic scheme (and used directly to test some aspects of the theory).

Starting at the bottom of the defence system with the periaqueductal grey, what we require is a stimulus maximally activating this region accompanied by minimal activation of other parts of the defence system. With such a challenge we could then test patients for the extent to which the periaqueductal grey itself is over-reactive, as opposed to being secondarily triggered by excessive activity elsewhere in the defence system. The periaqueductal grey, as noted earlier, controls 'fight/flight reactions to impending danger, pain, or asphyxia' (Graeff *et al.*, 1996). 'Danger' in any general sense could clearly produce widespread activation of the defence system before activating the periaqueductal grey. As a first step in developing a diagnostic test, therefore, we suggest mild asphyxia as a challenge. More subtle assessment could be necessary; and, indeed, it seems that panic disorder may be detectable from irregularities in respiratory rhythm and perhaps the response to respiratory challenge (e.g. Papp *et al.* 1995). From the present perspective, the optimum test would detect panic, not only in clinically anxious populations but also in the relatively rare cases (rare at least in the clinic) of panic disorder not accompanied by anxiety. Probably the best test would be to expose subjects to increasing levels of CO_2 and determine their *threshold* for panic, as opposed to the extent of any attack. As soon as panic is elicited, other parts of the defence system could contribute to the attack. Challenge with fixed levels of CO_2 is not only theoretically unattractive but does not discriminate panic well from, for example, specific phobias (Antony *et al.* 1997). Threshold measurements, on the other hand, should detect supersensitivity in the periaqueductal grey independent of other abnormalities in the defence system. There may also be relatively input-specific abnormalities of the periaqueductal grey whose detection would require testing with, say, painful stimuli as well as asphyxia.

At the next level up, we have the hypothalamus and at least some aspects of amygdalar function. If our current analysis is correct, no new test is required here, as the

only disorders of major concern are the specific phobias, which are simply diagnosed with current methods.

Then we come to the aspects of amygdalar function which we have associated with the arousal component of anxiety. The most obvious relevant challenge would be fear-potentiated startle, since this is not only sensitive to anxiolytic drugs (including when injected into the amygdala) but is also insensitive to hippocampal lesions. The only problem here would be if further work with animals were to show this test to be sensitive to hypothalamic or periaqueductal grey lesions (an issue which, to our knowledge, has not previously been investigated).

Next we come to the septo-hippocampal system. What is required is a test sensitive to septo-hippocampal system damage (or hyperactivity) and anti-anxiety drugs, but *not* to amygdalar or periaqueductal grey lesions (hypothalamic lesions would be likely to interrupt the theta frequency control system as well as escape and avoidance behaviour, and so their effects could not be used to rule out a proposed test). The most obvious tasks, here, are spatial navigation, delayed matching-to-sample, and behaviour on a fixed interval schedule of reward. Of these, delayed matching-to-sample can be most clearly set up in an anxiety-free form and so would probably be preferable, but it might be too specific in the aspects of septo-hippocampal function which it engages.

We have only limited clues as to what might constitute useful diagnostic tests for other anxiety-related disorders. There is 'selective, subtle evidence of autonomic dysregulation' in social phobics tested with autonomic challenges which did not include provocation of anxiety (Stein *et al.* 1994a, p. 218); but more work will be required to show that this dysregulation is not also present, for example, in panic disorder or agoraphobia. Similarly, tasks involving visual attention show abnormalities in obsessive–compulsive disorder patients (Nelson *et al.* 1993), but it will be necessary to show that this is not also the case in generalized anxiety disorder or as a simple consequence of anxiety. Tests of prefrontal hyperfunction could be based on the existing neuropsychological tests of prefrontal hypofunction.

11.21 CONCLUSION

Much work remains to be done before we have a complete neuropsychology of anxiety even for the rat, let alone people. However, we believe that this book provides a skeleton of the former; and we have, in this chapter, attempted to provide at least some suggestions for the latter. Most hopefully, the basic theory of the present edition is essentially the same as that of the first edition, while its degree of contact with the clinical literature on both anxiety and amnesia is greater. The theory is certainly not without flaws, probably multiple and major; but nonetheless the time has perhaps come to apply it relatively directly to clinical issues.

Some of the most significant changes that do differentiate the theory presented in this book from that of the first edition reflect the much greater knowledge that has accrued over the intervening years of the roles played by the primary defence system, including the periaqueductal grey, hypothalamus, and especially the amygdala. In the present chapter we have used this greater knowledge to try to pinpoint loci in the brain at which the

various symptoms and syndromes of anxiety and related states are mediated. This has led us (as have the specific considerations outlined in Section 11.3) into a relatively 'splitting' stance, emphasizing the different roles played in anxiety by different brain regions rather than what might unite them. In the final two chapters, which deal with new issues (personality and therapy, respectively), we shall find reason to reconsider this stance.

NOTE ADDED IN PROOF

Our hypothesis that hippocampal hyperfunction can lead to anxiety has been directly supported recently by a report (Crestani, F., Lorez, M., Baer, K., Essrich, C., Benke, D., Laurent, J.P., Belzung, C., Fritschy, J.-M., Lüscher, B., and Mohler, H. (1999) Decreased $GABA_A$-receptor clustering results in enhanced anxiety and a bias for threat cues. *Nature Neuroscience*, **2**, 833–9) that mice with impaired $GABA_A$-receptor function in the hippocampus, parahippocampus and orbitofrontal cortex show increased 'behavioural inhibition . . . and an explicit memory bias for threat cues'.

12 Putting Humpty Dumpty together again: the anxious personality and its inheritance

Many scientific controversies can be seen as debates between 'splitters' and 'lumpers': those who emphasize the differences between phenomena and those who see in them an underlying unity. The overall thrust of science takes the latter direction, to the point that a major effort in contemporary physics is to find a Grand Unifying Theory able to unite the already small number of fundamental physical forces into a single scheme of things. But there is no point in unifying where the data stubbornly indicate obstacles insurmountable to current concepts, or where emergent properties make a categorical approach convenient for a higher level of explanation. In the previous chapter (Section 11.3) we encountered obstacles of just this kind. Moreover, the main aim of that chapter—to align symptoms and syndromes with specific brain regions whose activity gives rise to them—inevitably, perhaps even tautologically, required of us a splitting mode of argument (albeit tempered by the high degree of recursive interaction linking the brain regions concerned). In this chapter, however, we consider a further area of data: the personality, and especially the heritable part of personality, of those who are particularly prone to the anxiety disorders. As we shall see, these more chronic predisposing factors, in contrast to the more acute factors which trigger specific symptoms, suggest an underlying commonality in the control, not just of different aspects of anxiety, but of the entire defence system. Similar inferences will emerge from our consideration of treatment issues in the following chapter. We shall therefore leave until the end of that chapter the necessary task of reconciling the splitting and lumping points of view.

12.1 THE ANXIOUS PERSONALITY

We do not all run equal risks of developing phobias, generalized anxiety disorder, obsessive–compulsive behaviour, and so on. Certain kinds of people are much more likely to manifest these symptoms than others. Analysis of just *which* people has, historically, been clouded by the very different approaches to this problem adopted by most psychiatrists, on the one hand, and most psychologists, on the other. The former apply a 'medical' model, as embodied indeed in the very term 'anxiety *disorder*'. Within this framework one seeks for specific causative agents and aetiological routes; and also for specific organs in which to locate primary pathological processes (an aspect of the medical model that we pursued assiduously in the previous chapter). Psychologists, in contrast, see many psychiatric disorders, and certainly those involving anxiety, as reflecting merely extreme positions on statistical distributions of behavioural propensity, positions that may fluctuate to some degree over time but are largely stable and so define an individual's personality. On this view, most psychiatric problems are not usefully regarded as analogous to

physical diseases, with the attendant dichotomy of 'well' or 'ill'. Rather, there are a number of such continuous distributions of behavioural propensity which run through the entire population, and those individuals who need psychiatric attention are simply located near the extreme pole of one or other of these 'dimensions' (Eysenck 1960) (like those who need medical attention because they lie at the extreme 'hypertension' position on the normal distribution of blood pressure). In this way, then, it is possible to distinguish different types of behavioural disorder by their location in the multidimensional space so created.

Note that the space used to locate psychiatric syndromes in this way is one of which the axes are dimensions of *normal* personality (H. J. and S. B. G. Eysenck 1969). This approach is, of course, highly congenial to the perspective on anxiety, central to this book, provided by ethology and animal learning theory, both of which treat anxious behaviour as a normal constituent of an animal's survival repertoire. From this point of view, then, the solution to the problem of psychiatric classification is inextricably bound up with the problem of how best to describe normal personality. Fortunately, there is now considerable agreement about the number and general nature of the major independent (i.e. uncorrelated) dimensions of personality that are needed to define the individual who is prone to anxiety, whether these are situated within the Eysencks' well-known three-dimensional space or its main rival, the 'Big Five' model of Costa and McCrae (Watson *et al.* 1994). There is also increasing agreement among psychiatrists (e.g. Cloninger *et al.* 1993) that the dimensional approach is appropriate to many psychiatric disorders. This change in the climate of psychiatric opinion is driven in part by clinical evidence that points to an underlying liability to disorder that operates dimensionally. For example, comparisons across populations indicate that the prevalence of cases is strongly correlated with mean symptom scores in the normal range, even when diagnosed cases are excluded (Rose and Day 1990); and there is a high frequency of subthreshold disorders (Goldberg and Huxley 1992). Risk factors also appear to operate dimensionally. For example, the correlations between 'minor psychiatric morbidity' (mostly reflecting mixed anxiety and depression) and such risk factors as geographical area, employment status, and age function dimensionally (Anderson *et al.* 1993). This clinical evidence for dimensionality is strongly reinforced by genetic evidence, as we shall see later.

Within the dimensional framework, symptoms of anxiety and related conditions are found predominantly in individuals whose personality lies in the neurotic introvert portion of the two-dimensional space described by Eysenck and Eysenck (1969). Moreover, when additional dimensions are taken into account, whether the third, Psychoticism, dimension of H. J. and S. B. G. Eysenck (1976) or the more complex five-dimensional space of Costa and McCrae (1985), this statement requires no further qualification. The personality dimension most closely related to susceptibility to anxiety is that of Neuroticism, with a smaller contribution from low (i.e. introverted) scores on the dimension of Extroversion. Direct measurement of the personality trait most closely related to high anxiety—'trait anxiety'—gives loadings of about 0.7 on Neuroticism and about –0.3 on Extroversion. Individuals who suffer from almost any of the whole range of anxiety-related disorders tend to have personalities which fall into this neurotic introvert region of personality space. The major exceptions to this rule are people with specific phobias,

e.g. of small animals. The personality of these individuals apparently fits no particular pattern, but can lie anywhere in personality space (Marks 1969). Such phobias perhaps reflect more or less universal childhood fears that in some people simply persist into adulthood. This view is consistent with the lack, usually, of any spillover from specific phobias into other areas of the person's life, and with the comparative ease with which they succumb to behavioural treatment (of a kind considered in the next chapter).

The role played by a predisposing personality is less obvious also in post-traumatic stress disorder. Given a sufficiently extreme event, there is a very high incidence of this reaction, likely to obscure any influence that personality might have. For example, as many as 30% of the fire-fighters of an Australian bush fire had chronic post-traumatic stress disorder 29 months later (McFarlane 1989). Even so, in this same sample (given their occupation, unlikely to have been selected so as to contain many extremely neurotic individuals), the level of Neuroticism was a better predictor of post-traumatic stress disorder than the degree of exposure to the disaster itself. However, given the long-lasting and pervasive nature of post-traumatic stress disorder, this condition may itself be legitimately regarded as involving a *change* in personality, one that takes the form of a persistent increase in the reactivity of the entire defence system (see the discussion of Adamec's and King's experiments in Section 11.14). In this way, then, severe environmental circumstances in adulthood can exacerbate or even mimic personality traits which, in most cases, are established much earlier in life.

12.2 NEUROTICISM VS. TRAIT ANXIETY

We have so far assumed that Eysenck is correct to locate two of his major dimensions of personality along axes defined by questionnaires of Neuroticism and Extroversion; and that, correspondingly, the correct description of the personality that predisposes to high levels of anxiety and a high probability of developing an anxiety disorder is 'neurotic introvert'. There is, however, an alternative to this position.

In the first place, since the overall defence system, as we have described it, functions in at least many respects as a unity (albeit one with many differentiated parts), it is natural to locate anxiety-prone individuals at the pole of a single personality dimension rather than in a quadrant bounded by two, as does Eysenck's 'neurotic introvert' location. We have considered this issue before (Gray 1970b) and proposed the rotation of Eysenck's axes shown in Figs 12.1 and 12.2. This preserves two orthogonal axes in the space defined by Eysenck's dimensions of E and N, but now has one of them run from his neurotic introvert to his stable extrovert quadrant. Schematically, one can represent this as a 45° rotation of Eysenck's axes, as in Fig. 12.1. But the fact that cingulectomy and prefrontal lesions reduce neuroticism more than introversion (Powell 1979), coupled with the observed correlations between the Eysenckian dimensions and scales intended directly to measure the trait of anxiety-proneness (Gray 1970b), suggests that a rotation which located the resulting dimension of 'trait Anxiety' (the negative diagonal in Fig. 12.1) closer to Eysenck's dimension of Neuroticism would be more appropriate (Gray 1981). The dimension thus rotated (Fig. 12.2) is close to trait anxiety as this is found in the work of Cattell (e.g. 1965) and Taylor (1953).

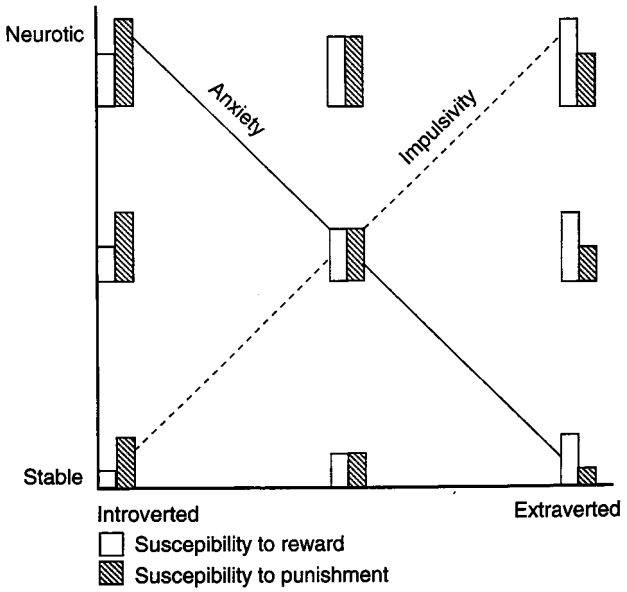


Fig. 12.1 A graphical representation of a simplified two-factor version of reinforcement sensitivity theory in which the fundamental personality axes (anxiety; impulsivity) are shown at a 45° angle to Eysenck's dimensions of extroversion (horizontal axis) and neuroticism (vertical axis). Although this figure is similar to one which has appeared in previous papers it includes an important correction (see text for details). (From Pickering *et al.* 1998.)

A rotation of this kind can be proposed because the multivariate statistical techniques of factor analysis, used by both the Eysencks (Eysenck and Eysenck 1969) and Cattell (1965) as well as many others, can establish *how many* factors or dimensions there are in a given space, but not *where* they should be located. Thus, if any location of these axes reflects underlying causal influences (and it is entirely possible that none does), this cannot be established by factor-analytic techniques alone but must be justified by other considerations, empirical or theoretical. The rotation of Eysenck's axes shown in Fig. 12.2 can be justified in part in terms of parsimony: we now have to suppose that the effective physical therapies for anxiety (drugs and prefrontal or cingulate lesions) shift individuals along only one dimension rather than two (Gray 1970b). Other reasons for making the rotation, germane more to personality theory than to anxiety as such, can be found in Gray (1970b, 1981) and Gray *et al.* (1983). Note that Fig. 12.1 (taken from Pickering *et al.* 1998) differs slightly, but importantly, from the original version (Gray 1970b, Fig. 7), which has been reproduced many times since, including in the first edition (as Fig. 16.1) of this book. The original erroneously illustrated Gray's (1970b) proposal as including an interaction between the Eysenckian dimensions of Neuroticism and Extroversion, in that the relative heights of the columns showing susceptibilities to 'signals of punishment' (reactivity to which determines position on the trait Anxiety diagonal) and 'signals of reward' (reactivity to which determines position on the orthogonal diagonal of Impulsivity), respectively, increased disproportionately fast as Neuroticism increased. The modified Fig. 12.1

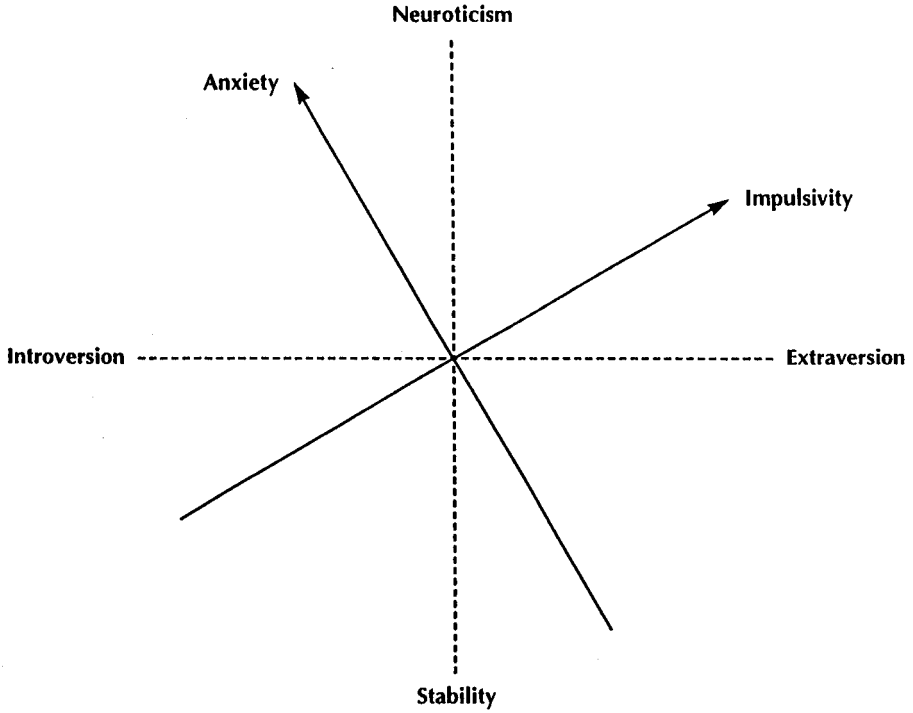


Fig. 12.2 A simplified geometrical representation of a more realistic set of relationships between the 'rival' personality axis systems of anxiety–impulsivity and extroversion–neuroticism. The axes lie at a 30° angle to one another, rather than the 45° angle shown in Fig. 12.1. The axis labels are at the high poles of the anxiety and impulsivity axes. (From Pickering *et al.* 1998.)

shown here illustrates correctly the chief postulates of Gray's (1970b) model. The two-dimensional space defined by Eysenck's factors of Neuroticism and Extroversion results from individual variation in the reactivity of two separate brain systems. Reactivity in one of these determines sensitivity to stimuli associated with punishment or frustrative non-reward. This determines trait Anxiety (which is 30° rotated from Neuroticism). Reactivity in the other brain system determines sensitivity to stimuli associated with reward or relieving non-punishment. This determines Impulsivity (which is orthogonal to anxiety proneness). (The Impulsivity aspect of Gray's model does not concern us here; for recent treatments, see Depue and Collins, 1999, and the commentaries thereon by Gray, 1999, and Pickering, 1999; and Pickering and Gray, 1999.) Eysenckian Neuroticism and Extroversion are then seen as derived dimensions, reflecting the (unequal; Fig. 12.2) mixtures of trait Anxiety and Impulsivity that give rise to them.

Having rotated Eysenck's axes in this way, we can now offer an essentially axiomatic answer to the question: what is the anxiety-prone personality? For the model shown in Figs. 12.1 and 12.2 states that such an individual is one who is highly susceptible to the adequate stimuli for anxiety (i.e. threats of punishment, of failure, innate anxiety stimuli, etc., as discussed earlier in the book). This statement, clearly, is perilously close

to circularity. But we are saved from circularity by the theoretical elaboration of the behavioural inhibition system that has gone before. This permits the derivation of many new predictions not contained in the description of an anxious person simply as one who is sensitive to anxiogenic stimuli. Furthermore, like any other aspect of personality, trait anxiety is a continuing disposition, present at all times; and scores on this dimension are continuously distributed in the normal population. It thus becomes possible to test predictions by seeking appropriate behavioural differences under laboratory conditions between normal individuals with differing levels of trait anxiety. This, of course, is the same strategy that Eysenck (1967), Cattell and Scheier (1961), and Spence and Spence (1966) employed in classic studies to test their several theories of anxiety. As we shall see, research of this kind suggests that anxiety-proneness is best thought of as a rather general trait, in a manner that is well captured by retaining the term Neuroticism for the dimension even after its 30° rotation from Eysenck's original placing of it. Far from being trivial, then, the link between anxiety and the anxiety-prone personality opens up the theory to experimental attack across a far wider range of data than would otherwise be possible. Indeed, so vast is the reach of these data that to measure the theory against them calls, not for a chapter, but another book.

A complication in the writing of such a book would arise from the changed definitions given here, compared with the first edition, of the adequate inputs to the behavioural inhibition system: to the stimulus-oriented description of these inputs employed previously, we have added the requirement that they must elicit conflict (as discussed at many points throughout the book). This potentially important distinction has not been at the forefront of the minds of the majority of investigators (including ourselves; e.g. Pickering *et al*. 1997) concerned with the experimental study of behaviour as a function of personality. With the significant exception of Newman's (e.g. 1987) work (emphasizing the different personality correlates of responses to punishment as such and those to punishment delivered in a context of reward, respectively), most such research has concentrated on a stimulus-oriented description of the conditions able to activate the behavioural inhibition system and, consequently, able to display behavioural differences between individuals with high and low scores on Neuroticism or trait anxiety. Given the greater elaboration of the different types of output from the separate components of the overall defence system, and the resulting more differentiated psychopathology (as discussed in detail in Chapter 11), to which the data now direct us, it is possible that we should see Neuroticism as relating to the whole gamut of anxiety-related disorders rather than to just anxiety itself. If one pursues this line of thought (for which, as we shall see, there are strong supportive data), one emerging possibility is that Neuroticism or trait anxiety reflects general sensitivity to threat (including the adequate stimuli for the behavioural inhibition system), irrespective of whether or not these give rise to anxiety (*ex hypothesi* the latter state is produced only if there is, in addition to the presence of threat-related stimuli, an element of conflict). This stimulus-oriented description of trait anxiety as reflecting the level of sensitivity to threat (in its broadest sense) is essentially the approach adopted in Gray's (1970b) initial formulation of the model illustrated in Figs 12.1 and 12.2. This approach raises, however, a terminological problem. If *trait anxiety* reflects sensitivity to threat irrespective of the occurrence of conflict, whereas *states of anxiety* require in addition the presence of conflict, then 'anxiety' in the two locutions has different meanings.

For the present we can do no more than hint at some of the themes that a book on the anxious personality should address. Many of the relevant observations have been reviewed by Eysenck (1957, 1967, 1981) and incorporated within his theory of personality. Although this theory has many striking successes to its credit, the strains imposed on it by a number of key experimental findings are probably too great for it to survive without modification (Gray 1981). But, like any good scientific theory, its demise will eventually come at the hands of a better theory, rather than the experimental findings alone. Whether the present theory, suitably extended into the field of personality, provides a viable alternative to Eysenck's cannot be determined without a thorough analysis of the existing data from the new vantage point it offers, and we cannot attempt this here. But, as argued elsewhere (Gray 1981), it seems in principle to be able to get round some of the problems with which Eysenck's theory is faced. Furthermore, there are certain areas of data for which the two theories make conflicting predictions and the present theory is upheld.

The most notable of these concerns Eysenck's (1957) conditioning postulate, that is, the hypothesis that introverts (and especially neurotic introverts) form conditioned reflexes more readily than extroverts. This postulate forms a critical link in the chain of argument by which Eysenck deduces that the neurotic introvert will be especially likely to manifest symptoms of anxiety (Gray 1970b; see Fig. 12.3). In contrast, the model presented in Fig. 12.1 postulates that introverts (and especially neurotic introverts) are particularly sensitive to secondary aversive stimuli, but relatively *in*sensitive to secondary positive reinforcers. Thus, if aversive conditioning is studied, both theories predict that introverts will outperform extroverts; but, if appetitive stimuli are used, Eysenck's theory predicts superior conditioning in introverts, whereas Gray's (1970b) predicts an extrovert superiority. The data lend support to the latter theory at this critical point (Gray 1970b, 1981; McCord and Wakefield 1981). The failure of Eysenck's conditioning postulate leaves him unable to predict the greater susceptibility to anxiety of the neurotic introvert (Gray 1970b); and this, from the present point of view, is the central problem in this area of personality theory.

The best way to determine the usefulness of a new approach to an old problem is usually to subject it to direct experimental test, rather than re-examine existing data. With regard to the model presented here, however, there is a major practical problem that must be borne in mind. The concept of the behavioural inhibition system is based on the behaviour of animals in situations in which there is little doubt that the animal's first concern is with the task that the experimenter has set it (to avoid a shock, to obtain food, etc.). From observations of this kind there has emerged a view of anxiety as something *useful*, a process that serves the adaptive function of steering the animal away from danger or eliminating useless responses. This is in stark contrast to the view of anxiety that has most often emerged from studies of human beings, whether in the psychiatric clinic or the experimental laboratory. Here, anxiety appears as a disruptive influence that prevents the subject from getting on with other things. In the laboratory, this situation probably arises because the tasks that are set a human subject are most unlikely, in the majority of cases, to become his first concern. While the experimenter wishes the subject to learn a list of nonsense syllables, his mind (if he is anxious) is on other things—am I being evaluated; will I pass my examination tomorrow; are there going to be electric shocks in this laboratory? This problem has been discussed by Mandler and Sarason

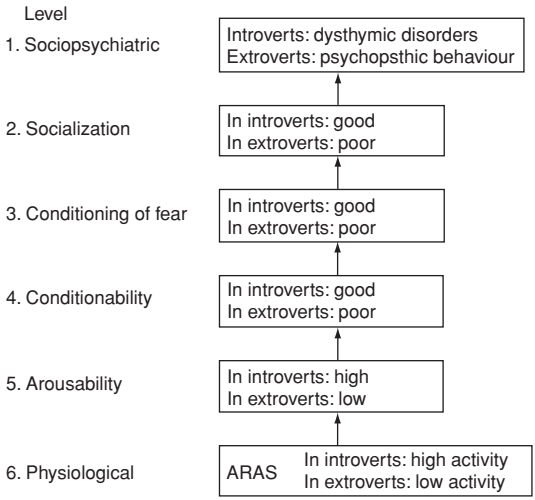


Fig. 12.3 The structure of Eysenck's (1967) theory of the development of dysthymic behaviour in neurotic introverts and psychopathic behaviour in neurotic extroverts. Each level is thought to give rise causally to the next level above. ARAS, ascending reticular activating system. (From Gray 1970b.)

(1952), who accordingly distinguish between task-relevant and self-oriented responses (see M. W. Eysenck 1977, Chapter 10). Many of the predictions that can be derived from the concept of the behavioural inhibition system are applicable only if the subject's behaviour is largely dominated by task-relevant concerns.

To take one example, it is clear from the description of the behavioural inhibition system given in this book that an individual with elevated sensitivity to frustrative non-reward (*ex hypothesi* one high in trait anxiety) should be better at reversal learning than an individual low in trait anxiety. This is the exact reverse of the hypothesis derived by Spence and Spence (1966) from a Hullian analysis of anxiety. These workers deduce that subjects high in trait anxiety should have particular difficulty in learning the correct response under conditions in which it is first necessary to inhibit a high-probability incorrect response to the same stimulus; and, by and large, the data support this deduction (Eysenck 1977, Chapter 10). But these data have been gained typically in paired-associate learning tasks; and it is reasonable to suppose, with Mandler and Sarason (1952), that such a task leaves plenty of emotional room for the intrusion of task-irrelevant concerns which, by competing for cognitive space, are likely to disrupt the performance of anxious individuals more than that of individuals low in anxiety. If anxious individuals are genuinely worse at reversal learning, it is strange that in animals anti-anxiety drugs impair the ability to suppress dominant incorrect responses and thus hinder reversal learning (Chapter 4). But the response the animal has to reverse is of overwhelming importance to its well-being. Thus the correct comparison is with the reversal of a response that is similarly important for human subjects and so likely to engage the behavioural inhibition system. It is not easy to see, however, how this can be achieved within the usual constraints of laboratory experiments.

Many similar predictions can easily be generated from the theory developed in this book, although usually hedged by the same practical qualifications. A more or less mechanical way of generating them is to make use of the effects of the anti-anxiety drugs described in Chapter 4: as a drugged animal is to a sober one, so an individual low on trait anxiety should be to one high on this trait. (The reverse of this argument, however, is not necessarily valid. If trait anxiety reflects general sensitivity to threat, irrespective of the presence of conflict, then there might well be differences between individuals high and low on this trait which are *not* paralleled by differences between animals given or not given an anxiolytic drug.) Other predictions can be derived from the more abstract description of the behavioural inhibition system given in the later chapters of this book (see especially Chapter 10). For example, the capacity for multidimensional stimulus scanning attributed to the behavioural inhibition system implies that anxious individuals should learn more about the different aspects of a set of stimuli than individuals low in anxiety—provided, as before, that they are anxious *about* the stimuli rather than attending to other things. When the first edition of this book was published in 1982, we noted that the scanty data relevant to this prediction were again negative. Thus, M. W. and M. C. Eysenck (1979) found that, with Neuroticism scores equated, introverts had greater difficulty than extroverts in scanning verbal material simultaneously for physical and semantic features. Again, however, there was no reason to suppose that the task was of a kind to engage the behavioural inhibition system. Since then, however, there has been a proliferation of studies of the attentional biases of both patients with currently high levels of anxiety and normal subjects high on measures of trait anxiety. These studies have abundantly confirmed the prediction we made in 1982 (although, unfortunately, usually without acknowledging its existence): there is a strong tendency for the attention of anxious individuals to be drawn to anxiogenic stimuli *provided that these are relevant to their concerns* (Mathews and MacLeod 1985, 1994; Eysenck and Byrne 1992; McNally 1996; Power and Dalgleish 1997, p. 230; and see Chapter 11, Section 11.10).

In this way, then, the model developed in 1982 and elaborated further here has received strong support. In other ways, however, the position is less clear. The difference between Neuroticism and an axis of trait Anxiety rotated 30° away from it (Fig. 12.2), as the principal predictor of susceptibility to the anxiety disorders, is probably too slight to be resolved with present epidemiological and clinical methods. (For this reason, and given also the potential confusion, noted above, between the meanings of 'anxiety' in the locutions 'trait' and 'state anxiety', respectively, we shall henceforth use the term 'Neuroticism', without further distinction as to its location, Eysenck's or a 30° rotation, to refer to individual differences in susceptibility to the anxiety-related disorders.) Experimentally, however, there has been a substantial and diversified effort to test Gray's (1970b) model of personality or subsequent elaborations of it (e.g. Fowles 1980; Gorenstein and Newman 1980; Newman 1987; Cloninger *et al.* 1993). Unfortunately, no clear picture has emerged from these studies (Pickering *et al.* 1997; Matthews and Gilliland 1999). Pickering *et al.* (1997) note the 'bewildering array of outcomes' of tests of Gray's (1970b) model and its congeners, and are tempted 'to abandon it altogether'. They go on, however, to remark that 'one cannot but remain impressed by the sheer frequency with which significant relationships nonetheless do emerge between one or other relevant personality trait and one or other relevant

change in behaviour due to reinforcement effects. Somewhere in the human brain there clearly are systems which influence individual differences in sensitivity to reinforcement' (Pickering *et al.* 1997, p. 63). In short, the jury is still out.

12.3 THE GENETICS OF NEUROTICISM AND EMOTIONALITY

Important for the argument followed here, the heritability of both Neuroticism and Extroversion is about 50% (Floderus-Myrhed *et al.* 1980), that is, about half the variance in scores is due to genetic factors inherited from parents. Since these scores, on each dimension, are distributed according to a normal Gaussian curve, the most likely underlying genetic architecture is one in which several polygenes, each of relatively small effect, contribute additively to the final phenotype. The anxiety-related disorders themselves are also substantially heritable (estimates range from 30 to 40% for general anxiety and phobic disorders; Kendler *et al.* 1992a,b). Critically, what appears to be inherited is the general predisposition to neurosis, rather than specific symptoms (Andrews *et al.* 1990). Thus, comorbidity of generalized anxiety disorder with other conditions, extending even to neurotic depression, is dependent on heritable factors to the same extent (30%) as generalized anxiety disorder itself (Kendler *et al.* 1992c). The remainder of the reliable variance in individual susceptibility to these conditions is largely due to specific within-family environmental factors (that is, life events that happen uniquely to the individual concerned), the between-family environment (including such socioeconomic factors as class, poverty, access to educational materials, etc.) playing only a minor role. The best single index of the heritable component of the comorbidity between the neurotic disorders consists in scores on Eysenck's Neuroticism scale or on similar scales highly correlated with it. These facts have major implications for the unity of the neuropsychology of anxiety and, indeed, for the unity of the whole range of defensive behaviour and its associated emotional states: despite the complexity of the hierarchical systems underlying active defence and passive avoidance, as considered throughout this book, there is heritable control over a single, quantitatively varying susceptibility towards suffering from any or all of the neurotic disorders, be they termed panic, anxiety, or even depression; and this is so irrespective of which particular brain mechanism proximately mediates the symptoms displayed (as considered in detail in the previous chapter).

Recent developments in allying the methods of statistical and molecular genetics have opened the way to the localization, identification, and eventual cloning of the additive polygenes which, as noted above, are likely to underlie the heritability of quantitative traits such as Neuroticism. The first step is to identify the chromosomal region in which the polygenes are located, using methods of 'quantitative trait locus' (QTL) analysis (Plomin *et al.* 1994). This approach is based upon the availability of polymorphic DNA markers (e.g. dinucleotide tandem repeats of variable length) mapped fairly evenly and densely across each chromosome. Such maps are now available for both rodents (the mouse map being somewhat ahead of the rat one) and human beings. Using maps of this kind one can seek for associations between each marker and scores on a quantitatively varying phenotype, e.g. one measured behaviourally. Multivariate statistical packages

are available that allow one to seek such associations simultaneously for many markers, and to interpolate to the most probable location of a polygene within a QTL region defined by flanking markers (e.g. Fulker *et al.* 1995).

We (Flint *et al.* 1995) have applied this strategy to what we believe is a rodent analogue of Neuroticism or trait anxiety. There is a long history of experimental tests of such a trait in animals, and evidence that behaviour in these tests is both related to the human trait of Neuroticism and substantially heritable (Broadhurst 1960; Gray 1987b). Research aimed specifically at measuring this trait (often termed 'emotionality') and determining its heritability started with Hall's (1951) selective breeding studies in the rat in the 1930s, replicated by Broadhurst at the Institute of Psychiatry (Maudsley Hospital), London, in the 1960s (reviewed by Gray 1987b). The criterion for selection in London was defecation (number of faecal boli) in an open field, under bright lights and white noise, high-scorers being brother–sister mated with high scorers, and low with low, over successive generations to produce two inbred lines, termed the Maudsley Reactive (MR) and Non-reactive (MNR) strains, respectively. Choice of defecation as a measure of fear was based on human experience, as formally documented in World War II, in which the strain of battle resulted in involuntary excretion in 21% of the sample studied (Broadhurst 1960). It will be important for the conclusions we reach later that, on our terms, defecation is a measure of fear rather than anxiety, since it is not affected by anxiolytic drugs.

Following the work of Hall and Broadhurst, DeFries initiated a replicated selection study of open-field activity in the mouse (DeFries *et al.* 1970). DeFries chose to select for high and low activity, rather than defecation, since activity is continuously distributed and more highly heritable than defecation, but the two are highly negatively correlated. DeFries and Hegman (1970), indeed, estimated the genetic correlation between these measures to be –0.86. The foundation population for DeFries' selection experiment was derived from an initial cross of two highly inbred strains (BALB/cJ and C57BL/6J). Beginning in 1966, six closed lines were formed: two were selected for high open-field activity (H_1 and H_2); two for low open-field activity (L_1 and L_2); and two randomly mated within line to serve as unselected controls (C_1 and C_2). Within-litter selection was practised, and the number of mating pairs per line was 10 or less per generation. The parents for each generation were selected by identifying the highest activity male and female from each litter for the high lines, and lowest activity animals for the low lines. These were randomly mated within lines. This procedure minimizes the possibility of chance fixation of alleles. Thirty generations of selection were completed, and a marked and highly reliable response to selection was obtained. In generation 30, the open-field activity scores of H_1 and H_2 were about 30 times higher than those of L_1 and L_2, on average, and the distributions of the high- and low-selected lines were non-overlapping. In 1976, selection was suspended and the six lines were random-mated (within line) for 18 generations. Subsequently, the lines were inbred employing brother–sister matings, and more than 35 generations of full-sib mating have been accomplished to date.

By generation 30 the open-field defecation scores of L_1 and L_2 were approximately seven times higher than those of H_1 and H_2, substantiating previous evidence for a large negative genetic correlation (about –0.8) between these two measures. Thus, just as selection for open-field defecation produced a correlated change in activity in the Maudsley rat strains (see below), selection for activity produced a correlated change in defecation

in the DeFries mouse strains. This result confirms other evidence of the considerable commonalities between rats and mice in open-field and related emotional behaviour (Royce 1977). Differences between the high- and low-selected lines were also observed when they were tested in types of apparatus (an arena, a barrier, a hole-in-the-wall) that are box-like, brightly illuminated, and somewhat resemble the open field (De Fries *et al.* 1970). However, these differences did not generalize to more confining situations like the Y-maze or staircase, or to performance in exercise wheels. Thus, the DeFries high- and low-selected lines differ in emotional reactivity to novel test situations similar to the open field, but not in measures of general activity level. We further asked whether these differences in emotional reactivity extend beyond the open-field test, by studying the H_1 and L_1 mouse strains for the first time on two behavioural tests (the elevated plus maze and the black–white exploration box) known to be sensitive to anxiolytic drug action (Pellow and File 1986; Hendrie *et al.* 1993). As predicted, relative to the low-active mice, high-active ones behaved on these tests as though they had received an anxiolytic drug (increased time on the open arms of the plus maze and in the white portion of the black–white box). Thus, the DeFries strains appear to provide an excellent starting point for a QTL analysis of rodent open-field behaviour.

The Maudsley strains of rats have been maintained continuously (for over 35 years) since their selection (Broadhurst 1960) and have been intensively studied. They constitute the best existing rat model of Neuroticism with high heritability. The size of the difference between the strain means, as measured by *t*-test, is currently about $t = 6–7$ each generation. The strain difference in open-field defecation is not a simple metabolic effect, since the difference in home cage defecation takes the opposite direction (MNR defecating more; Broadhurst 1975). Broadhurst ruled out effects of both post-natal (by cross-fostering) and pre-natal (by reciprocal cross) maternal environments as factors in the observed strain differences. He also reported a correlated response to selection on activity measures in the open field, MR being less active than MNR. This difference, too, has remained reliable across 35 years ($t = 7$). The persistence over the years of these strain differences is in spite of vicissitudes in the breeding program, and holds up also across sub-lines of the original Maudsley populations, based upon founders exported at various times from London (Blizard 1981). Broadhurst (1960) derived estimates of the narrow heritability of open-field defecation and activity from a 6×6 diallel cross, including the MR and MNR strains; these were 46 and 68% respectively. These figures are, however, somewhat inflated since they refer to the heritability between litter means; inspection of Broadhurst's published results suggests true heritabilities nearer one-half these figures.

The Maudsley strains have been shown to differ on many behavioural and physiological parameters over the years (for reviews, see Broadhurst 1975, who lists 280 separate observations; Blizard 1981; Gray 1987b). The range of these differences may have occurred by chance because the method used, selective inbreeding, should have resulted in both selected and chance fixation of alleles. Physiologically, selection has led to differences affecting gonadal and thyroid function, as well as responses to a wide range of drugs. Much of the behavioural evidence suggests that the MR and MNR lines differ for a broad range of emotional reactivity. Thus, relative to the MNR strain, the high-defecating and low-active MR animals explore less (take longer to emerge from a familiar to a novel

environment in an emergence test, show less rearing behaviour), show higher heart rate, especially in response to stress, have lower food and water intake under threat, show a greater frustration effect to food omission in Amsel and Roussel's double runway, greater conditioned suppression of bar pressing, greater passive avoidance, faster escape from shock and from immersion in water, and poorer shuttlebox avoidance (Broadhurst 1975; Gray 1987b).

Interpretation of these findings as indicating greater fearfulness in the MR strain is relatively straightforward (Gray 1987b), except perhaps for the MR inferiority in shuttlebox avoidance. Wilcock and Fulker (1973), however, found that two separate genetic factors affect performance in this task, both with directional dominance, one favouring good shuttlebox performance late in acquisition, the other favouring poor performance early in acquisition. The latter factor reflects the interfering effects of passive avoidance (of the previously shocked side) in this complex task (Gray 1987b; and see the discussions of the effects on shuttlebox avoidance of anxiolytic drugs in Chapter 4 and of septal and hippocampal lesions in Appendix 8). More specific interpretation of the overall pattern of findings as being related to human anxiety is suggested by parallels with the effects of anxiolytic drugs (especially benzodiazepines). As reviewed in Chapter 4, these generally act in unselected animals to reduce behaviour high in the MR strain: they reduce emergence time, increase rearing behaviour, decrease conditioned suppression and passive avoidance, and improve shuttlebox avoidance. However—and critically important for the argument pursued here—the differences between the Maudsley strains also extend to features not paralleled by the profile of action of anxiolytics, notably open-field defecation and activity themselves, escape from shock, and the double-runway frustration effect. Furthermore, echoing the data on human Neuroticism considered above, these differences extend even to an animal model of depression, the forced swim test, in which MR rats obtained scores indicative of higher susceptibility to this condition (Commissaris *et al.* 1996).

The overall pattern of differences between the Maudsley strains is, thus, consistent with the hypothesis that they differ in sensitivity to the overall effects of threatening stimuli. This difference would then alter input to, and hence affect the functioning of, *both* the behavioural inhibition system *and* the fight–flight system (Gray 1987b, pp. 266–8); or, in terms of neural structures, both the septo-hippocampal system and those that make up the primary defence system (see Chapter 6). The involvement of the behavioural inhibition system is supported by observations of strain differences in the hippocampal theta response to septal driving stimulation (MNR rats resemble unselected rats treated with anxiolytic drugs; Drewett *et al.* 1977) and in the functioning of the dorsal ascending noradrenergic bundle (Blizard and Liang 1979; Sara *et al.* 1993; Verbanac *et al.* 1994). The involvement of the fight–flight system is suggested by the strain differences in autonomic reactivity (defecation, heart rate), shock escape, and the double-runway frustration effect; thus, for example, defecation is reduced by lesions of the central grey (Liebman *et al.* 1970) and the frustration effect by lesions of the amygdala (Henke 1977), but neither of these changes is seen after damage to the septo-hippocampal system (Appendix 8). The Maudsley strains differ also in the concentration of serotonin in the limbic system (Maas 1963); this may affect processing in both the behavioural inhibition and the fight–flight system (Chapter 11). However, these strains do not differ in the activity of dorsal raphe neurons (Verbanac *et al.* 1995).

The DeFries mouse and Broadhurst rat selective breeding programmes have shown that, starting from different behavioural selection criteria (open-field activity and defecation, respectively), one can nonetheless arrive at very similar behavioural phenotypes: each programme affected a number of additional forms of behaviour, including the alternative selection criterion. A similar conclusion emerges from breeding programmes based upon the selection criterion of performance in a shuttlebox avoidance task (Bignami 1965; Brush *et al.* 1985). These have led to correlated changes in open-field defecation and locomotor activity that mirror those found in the breeding of the Maudsley rat strains. In both cases, poor shuttlebox performance turns out to be associated with high defecation and low activity. These data reinforce the impression of a unified system, manifest in a variety of tests of emotional behaviour, some (e.g. shuttlebox avoidance) sensitive to anxiolytic drugs and some (open-field defecation) not.

Given the greater density with which the mouse, relative to the rat, genome has been mapped so far, we commenced our rodent QTL studies with the DeFries mouse strains. One pair of the high- and low-active strains was crossed to produce an 'F_1' population (each of the offspring—genetically identical because of the highly inbred status of the parental strains—receiving half its genes from one parent from each strain). This population was then randomly mated to produce 879 'F_2' intercross animals (a breeding régime which effectively 'shuffles' the genes from the two parental strains to make new mixtures in each of the offspring). These animals were subjected to a battery of behavioural tests and then genotyped. The tests used consisted of activity and defecation in the open field, activity in a Y-maze, and open- and closed-arm entry in the elevated plus maze. The results are summarized in Fig. 12.4 (Flint *et al.* 1995), based upon statistical analysis (Lander and Botstein 1989; Lincoln *et al.* 1992) of the observed associations between each behavioural measure and specific, chromosomally localized, DNA markers that were dimorphic in the parental H_1 and L_1 strains. The results for chromosome 1 are particularly striking and clearly suggest that a single pleiotropic QTL underlies four of the five measures. The fifth measure, entry into the closed arms of the elevated plus maze, was included as a negative control, since (unlike entries into the open arms) it is not affected by anxiolytic drugs. Thus its failure to relate to the chromosomal 1 QTL is as predicted. Since this work was published, a QTL for emotionality on chromosome 1 has been replicated independently (Caldarone *et al.* 1997; Gershenfeld *et al.* 1997). The results for chromosomes 12 and 15 were also consistent with similar QTLs, but they were somewhat less striking; it is possible that on these chromosomes there are linked multiple QTLs that our data failed to resolve. Between them, however, the QTLs detected account for essentially all the heritable variance in open-field behaviour in these strains (Flint *et al.* 1995). Thus it would appear that this experiment succeeded in detecting all the polygenes responsible for the divergence of the DeFries mouse strains on the behavioural test used selectively to breed them. Other selection experiments may very well have engaged other genes. Nonetheless, it is encouraging for this research approach that, at least in the case of the DeFries mouse strains, a tractably small number of polygenes is capable of determining variation on a continuous quantitative trait analogous to human Neuroticism. Furthermore, the pleiotropy detected in this experiment for the QTL on chromosome 1 affected behaviour that both is (entry into the open arms of the elevated plus maze) and is not (open-field defecation) affected by anxiolytic

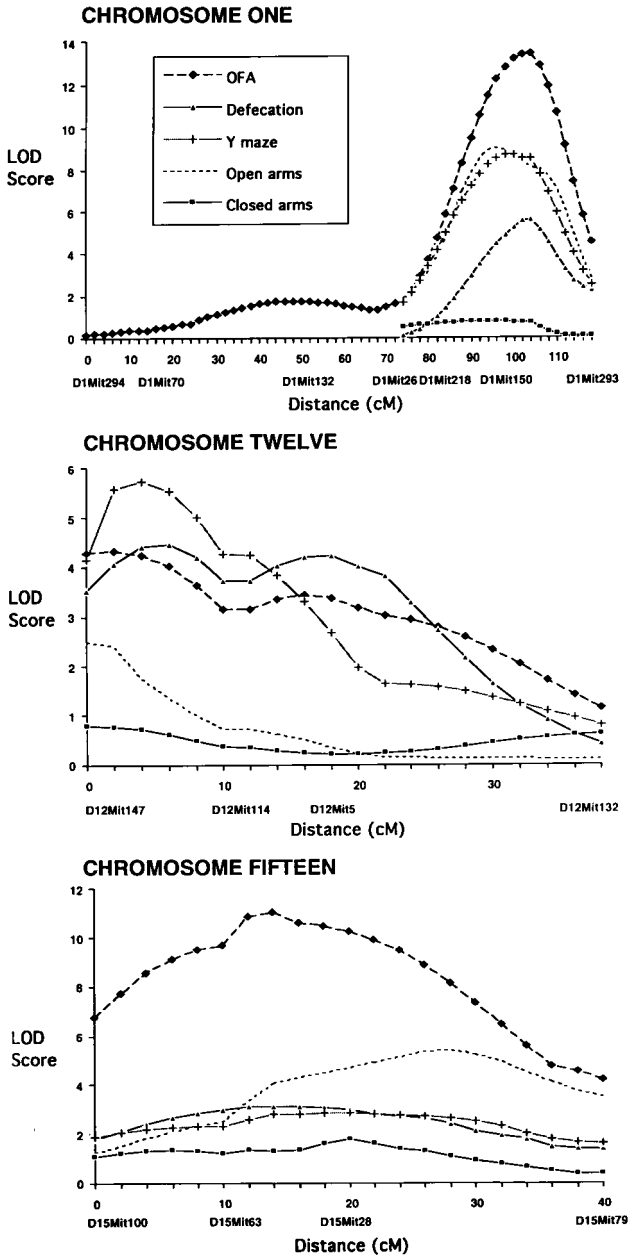


Fig. 12.4 Quantitative trait locus (QTL) analysis of five measures (open-field activity [OFA], defecation in the open field, activity in the Y-maze, and entry into the open and closed arms of the elevated plus maze) on chromosomes 1, 12, and 15. The LOD score curves were generated by the MAPMAKER-QTL program (Lincoln *et al.* 1992). The approximate position of markers used is given below the *x* axis; cM, centimorgans. LOD, logarithm to base 10 of the probability of a chance observation. (From Flint *et al.* 1995.)

drugs (Chapter 4), suggesting therefore that operating characteristics of both the behavioural inhibition system and the fight–flight system are determined by the polygenes concerned.

These genetic data from studies of the inheritance of both human neurotic disorders and rodent emotionality strongly suggest, then, that what is inherited is a broad propensity to display a variety of forms of emotional behaviour which other considerations in this book lead us to believe require for their discharge a widely distributed array of brain mechanisms, spanning both the behavioural inhibition and the fight–flight systems. The additive polygenes that give rise to this pattern of inheritance must, therefore, specify some, as yet unknown, equally broad, operating characteristics or parameters of this distributed array. Such operating characteristics might include, for example, the level of sensitivity to the detection of threat by all levels of the defence system. This would be in line with the suggestion, above, that such sensitivity to threat underlies the degree of Neuroticism.

How unity is in this way forged from diversity, we can at present only guess. Some of the genes concerned may act to specify critical features of the activity of the recursive loops whose functions we have many times stressed throughout this text, and whose possible clinical effects were elaborated in the previous chapter. If the centrality that our theory offers to the septo-hippocampal system is well founded, then another possibility is that the polygenes specify critical operating characteristics in this system. This possibility is strengthened by a number of reports of correlations between genetic variation in the detailed wiring of the rodent hippocampus, on the one hand, and performance on behavioural tests related to both spatial cognition and emotionality, on the other (van Daal *et al.* 1991a,b; Schwegler and Crusio 1995; Sluyter *et al.* 1998, and earlier references therein); tests of the latter kind have included both open-field exploration (Crusio and Schwegler 1987) and shuttlebox avoidance (Schwegler and Lipp 1983). These data suggest pleiotropic genetic effects, mediated by specification of detailed hippocampal neuroanatomical parameters, that unite aspects of cognitive and emotional function in a manner broadly agreeing with the type of computational processing attributed by our model to the hippocampus.

Another possibility is that some of the polygenes concerned specify operating characteristics of the ascending monoaminergic projections which provide important biasing and modulatory inputs to the septo-hippocampal system, as well as to other components of the overall defence system. Genes specifying operating characteristics of both the noradrenergic and serotonergic systems could play this type of role. As we have seen, the Maudsley strains differ in features of both these systems (Maas 1963; Blizard and Liang 1979; Sara *et al.* 1993). The efficacy in a number of anxiety-related disorders of drugs that act on serotonergic transmission (see Chapter 11) make this route of genetic action particularly plausible. A final possibility is genetic specification of aspects of the hypothalamo–pituitary–adrenal axis. Transgenic mice developed to overproduce corticotropin-releasing factor show, along with elevated levels of adrenocorticotropic hormone and corticosterone, increased anxiety as measured in the elevated plus maze and open field (Stenzel-Poore *et al.* 1994; see also Bitran *et al.* 1998), two of the tests used in the Flint *et al.* (1995) study of QTLs in the DeFries mice. Conversely, mice lacking the corticotropin-releasing factor receptor CRFR1 not only lack the normal hormonal

response to stress but also show markedly reduced anxiety as measured in the light–dark box and elevated plus maze (Smith *et al.* 1998).

These various possibilities are not mutually exclusive. Recall that we are dealing with polygenes that have independent, additive effects. Thus we should expect that the functions influenced by these genes will, similarly, be additive and capable of independent effect (which, however, does not imply absence of interaction). The entire field of genetics is evolving at an extraordinarily rapid pace. It will not be long before the QTLs for emotional behaviour detected in the mouse cede place to the actual polygenes located there. Similar searches are under way for QTLs in the rat that may be syntenous (that is, located on the same stretch of DNA, as indicated by common markers) with those established in the mouse. Our own work focuses on the Maudsley and Roman rat strains; the latter have been bred (Bignami 1965) for high vs. low avoidance in the shuttlebox. If this line of research is successful, the greater store of physiological and behavioural knowledge available for the rat, relative to the mouse, will facilitate establishment of the exact functions discharged by the identified polygenes. Investigation has also commenced in several major centres of the QTLs for human Neuroticism. These studies follow essentially the same logic as that used by Flint *et al.* (1995; see above) for the mouse, but the greater variability of unselected human genotypes requires much greater sample sizes: up from 1000 rodents to about 20 000 pairs of siblings (Eaves and Meyer 1994; D. W. Fulker, personal communication). Should QTLs emerge from these studies that are syntenous with those established in rodents, it should be possible directly to test the functions of the polygenes concerned. Thus the speculations in which we have indulged in the preceding paragraphs, as to possible physiological functions of these polygenes, are very likely soon to come home to roost.

The data reviewed above, concerning the neurotic personality and its genetic basis, offer strong support for a more unified understanding of the neuropsychology of anxiety, and of defensive behaviour more generally, than was apparent in the more symptom-oriented approach that we adopted in Chapter 11. This understanding is consistent with the hierarchichal view of a differentiation of specific anxiety disorders within an overall 'trait diathesis common to all anxiety disorders' advocated by Zinbarg and Barlow (1996, p. 181; see also Brown *et al.* 1998). As we shall see, support for an integrated hierarchy of this kind is also found in the effects of psychological treatments for the anxiety-related disorders, a topic to which we turn in the next, and final, chapter.

13 The treatment of anxiety

The evidence reviewed in the previous chapter, dealing with personality and the inheritance of susceptibility to the anxiety-related disorders, began to redress the balance from the 'splitting' stance (to which we were driven by the data on symptoms and their neural mediation reviewed in Chapter 11) towards a more 'lumping' stance, one that is characteristic also of current clinical classifications of these disorders. Our conclusion in that chapter was that possession of a neurotic (trait anxious) personality increases reactivity to threat, and so creates a predisposition to the whole range of threat-related disorders (including anxiety but extending even to neurotic depression). In this final chapter, we shall find congruent conclusions emerging from data on the treatment of anxiety-related disorders.

13.1 BEHAVIOUR THERAPY

In the discussion of the specific phobias in Chapter 11 (Section 11.13) we largely discounted any role for classical conditioning in the development of these responses. Anyone familiar with the successes of behaviour therapy in the treatment of phobias (e.g. Mathews 1978) and the kinds of theoretical account that have been given for these successes (e.g. Eysenck and Rachman 1965) could be forgiven a certain sense of bewilderment at this conclusion. For these successes have been gained by treating phobias as though they *were* conditioned reactions and then subjecting them to extinction or counter-conditioning. If this assumption is wrong, then what is the efficacy of behaviour therapy due to?

First applied to specific phobias, for example of snakes, spiders, or heights, behaviour therapy has subsequently shown efficacy in much more complex and disabling conditions, including agoraphobia (Mathews *et al.* 1981; Marks and Marks 1990) and post-traumatic stress disorder (Jaycox and Foa 1998), for which the principal therapeutic ingredient is simple exposure (see below); and obsessive–compulsive disorder, in which exposure is combined with 'response prevention', that is, prevention of the compulsive rituals, leading to a surge of anxiety which is then able to extinguish (Baer and Minichiello 1990). The evidence suggests, however, that the efficacy of behaviour therapy has rather little to do with the theories on which the therapeutic methods were originally based: details of therapeutic procedure that these theories would suggest to be of critical importance—for example, the ordering of the sequence of presentation of phobic items, the presence or absence of relaxation after presentation of an item—turn out to play an insignificant role, if any. All that seems to matter (to a first approximation) is the total amount of time for which the patient is exposed to the phobic stimulus—the greater the exposure, the greater the therapeutic effect (Gelder *et al.* 1973; Marks 1973; Levis and Hare 1977; Teasdale 1977). Now this poses something of a dilemma. If one gets

better by being exposed to the phobic stimulus, why did one get ill by being exposed to it in the first place? And why does one not get better in the natural course of exposure in the real world, without the help of a therapist?

An answer to these questions which is consistent with the theory developed in this book was proposed by Lader (Lader and Wing 1966; Lader and Mathews 1968) and later elaborated by Watts (1971, 1979). According to these workers, the common element in methods of behaviour therapy which are superficially very different from one another is that they allow responses to the phobic stimulus to habituate. This view fits naturally with the arguments advanced in Chapter 11 (see the discussion of Öhman's experiments in Section 11.13.1). If much of the behaviour of phobic patients consists of innate reactions to stimuli to which they are particularly sensitive, it follows that the disappearance of such reactions is perhaps due to habituation of the kind described by Sokolov (1960; Horn and Hinde 1971). If habituation underlies the effects of behaviour therapy, the most important variable would be expected to be—as it seems in fact to be—total exposure time.

Before examining this view in closer detail, let us pause to consider what is involved in two of the methods of behaviour therapy most often used with phobic patients: systematic desensitization and flooding. We shall take extreme versions of the two methods, so as to contrast them more effectively. The descriptions we give are based on the way in which these two treatments were initially introduced into clinical practice, that is, by asking the patient to imagine the phobic stimuli. Subsequently, these 'imaginal' methods of therapy were replaced, for the most part, by '*in vivo*' methods, involving exposure to real rather than imagined phobic stimuli. The generic name for both types of treatment now is, therefore, 'exposure therapy'.

In systematic desensitization, stimuli are first graded into a hierarchy according to their capacity to elicit fear. The subject is then asked to imagine them in a sequence which corresponds to gradually increasing phobic power. Each stimulus presentation is typically short, since it is terminated as soon as the patient signals the presence of any anxiety. Immediately after the stimulus is terminated, the patient is instructed to relax deeply, using techniques in which training has been given before therapy commences. This, essentially, is the method developed by Wolpe (1958). It is based on the notion of 'reciprocal inhibition' of anxiety by relaxation: 'if a response antagonistic to anxiety can be made to occur in the presence of anxiety-evoking stimuli so that it is accompanied by a complete or partial suppression of the anxiety responses, the bond between these stimuli and the anxiety responses will be weakened' (Wolpe 1958). This is essentially the same concept as that of 'counter-conditioning', used by Amsel (1962) to account for phenomena such as the partial reinforcement extinction and partial punishment effects (Dickinson and Pearce 1977; Gray 1987b, Chapter 10). Relaxation serves the same function, theoretically, as food reward in these paradigms. The graded hierarchy of phobic power and the short stimulus presentation are intended to keep the level of anxiety down so that counter-conditioning is facilitated.

In flooding or implosive therapy (Stampfl 1970; Levis and Hare 1977), the patient is again asked to imagine aversive stimuli. But now the therapist attempts to maximize emotional arousal by describing the stimuli as vividly as possible and by opposing any attempt on the part of the patient to elude them. Apart from the fact that this all occurs

in imagination, it is tantamount to throwing a child who is frightened of water into the deep end of a swimming pool. In this respect flooding is diametrically opposed to the step-by-step gradualism of systematic desensitization. Like Wolpe's (1958) technique, however, it is based on the assumption that phobias are conditioned fears, and its purpose is to eliminate them. This is supposed to occur by simple extinction: 'the presentation of fear cues is expected to elicit a strong emotional response at first, but with repetition the emotional responding should subside' (Levis and Hare 1977). From this point of view, the brief and mild elicitation of fear on which systematic desensitization is based would be a slow and inefficient way to produce extinction.

Given the very different procedures used in these two therapies and the different theoretical analyses on which they are based, it would be reasonable to expect that if one worked well, the other would not. Quite the opposite is the case: they both work fairly well, and attempts to discriminate between their therapeutic effects have by and large failed (Levis and Hare 1977; Teasdale 1977; Mathews 1978). This suggests that neither of them works quite in the way it is thought to do.

The habituation model of the treatment of phobias proposed by Lader and Mathews (1968) was intended to apply only to the method of systematic desensitization, since the flooding technique was not yet widely known. According to this model, anxiety is primarily a state of over-arousal in the central nervous system with consequently high levels of activity in the autonomic nervous system. This view is close to the one advocated here, if we substitute 'overactivity in the behavioural inhibition system' for over-arousal. This substitution is not difficult to make: arousal is one of the functions we have attributed to the behavioural inhibition system; and the level of skin conductance, taken by Lader and Mathews (1968) as a measure of arousal, appears to be under at least partial control by the behavioural inhibition system (Fowles 1980; and see the discussion of Öhman's experiments in Section 11.13.1). The theory proposed by Lader and Mathews (1968) goes on to treat systematic desensitization as 'habituation occurring when the rate of habituation is maximal; that is, when the level of arousal is as low as possible consistent with clear consciousness'. On this view, the role of relaxation and the presentation of stimuli low in the phobic hierarchy for only short periods is to maintain arousal level as low as possible, this being thought to facilitate habituation of the phobic response. Theoretically, this approach is closer to the arguments used by proponents of flooding, since it is difficult to distinguish between the concept of habituation as used by Lader and Mathews (1968) and that of extinction as used by Levis and Hare (1977). In practice, however, the method that Lader and Mathews (1968) argue for is systematic desensitization. This is probably because, at the time, this was the only method in wide and successful use. This is a warning against taking too seriously the *post hoc* application of a theory to the known facts: an exercise in which this chapter is, of course, engaged.

Watts (1979) returned to the theme of habituation. He discarded the 'maximal habituation' model of Lader and Mathews (1968) on the basis of both clinical evidence and advances in understanding of the process of habituation. The hypothesis he proposed himself, is, however, a natural extension of Lader and Mathews'. He follows Groves and Thompson (1970) in supposing that the reactions to a repeatedly presented stimulus are determined by the interaction of two processes, one decremental (habituation, properly

speaking), the other incremental (sensitization). The net effect observed on any given trial then depends on the properties attributed to these two processes. Habituation is seen as relatively specific to the particular response elicited by a particular stimulus, and to depend on the formation of a 'neural model of the stimulus' after the manner proposed by Sokolov (1960). It is independent of stimulus intensity and grows with repeated training sessions. Sensitization, by contrast, is non-specific, affecting the general level of responsiveness only; it is positively related to stimulus intensity; and it at first grows, but then decays, especially over repeated sessions. In the short term (that is, immediately after a particular stimulus presentation), both habituation and sensitization are thought to decay over time, but sensitization decays more rapidly.

Watts (1979) specifically relates this model to the behavioural inhibition system as described by Gray (1975, 1976). As he points out, this system is activated both by novel stimuli (to which the concepts of habituation and sensitization are applicable) and by secondary aversive (including phobic) stimuli. Thus it becomes parsimonious to treat the loss of reaction to phobic stimuli during behaviour therapy as a species of habituation. The more completely developed model of the behavioural inhibition system presented in this book maps even better onto Watts' (1979) hypothesis. As we have seen, this model includes both a septo-hippocampal circuit for habituation and an ascending noradrenergic pathway of arousal (Chapter 10 and Appendix 10). We may, therefore, think of the decremental process in Watts' theory as taking place in the former circuit (but perhaps also in the amygdala, which, as we saw in Chapter 11, probably plays a nodal role in the specific phobias) and the incremental process, in the latter.

We still have, however, separate counter-conditioning and habituation theories of behaviour therapy: how is one to explain the equal success of two such different therapies as systematic desensitization and flooding? Watts' (1979) article also opens up a useful line of attack on this problem.

He argues that two different combinations of conditions can be derived from his model as being therapeutically effective. On the one hand, one can attempt to minimize the effects of sensitization, while allowing habituation to exert its beneficial effects. For this one would choose stimuli of low phobic power and present them for short enough periods so that sensitization does not occur to any great extent. This, of course, is the method of systematic desensitization. Relaxation, as in Lader and Mathews' (1968) treatment, is also seen as preventing sensitization (arousal increment). In addition, since habituation but not sensitization is stimulus specific, it should aid therapy if the stimulus is clearly perceived. In support of this deduction, Watts cites an experiment of his own (Watts 1974) which showed that a careful description of the stimulus, each time the subject was instructed to imagine it, enhanced the amount of long-term reduction in anxiety reduced by systematic desensitization. (Note the congruence between this argument and the function of multidimensional stimulus analysis, attributed to the septo-hippocampal system in Chapter 10.) In contrast to this combination of conditions found in systematic desensitization, Watts (1979) proposes that flooding depends for its therapeutic success on the fact that, although it maximizes sensitization, it also (by keeping the subject exposed to the phobic stimulus for long periods) allows sensitization time to decay.

From this theory Watts (1979) is able to derive a number of specific predictions which palliate a *tour de force* that is otherwise disturbingly *post hoc*. Some of these predictions

seem to be supported by existing data. Thus, for flooding to work, long sessions should be essential, a deduction supported by the observations of Stern and Marks (1973). Conversely, the optimal conditions for desensitization should be short stimulus presentations together with relaxation. In accordance with this deduction, Proctor (1969) found 5-second stimulus presentations to be superior to 20-second presentations if the subject was required to relax; but without relaxation the relationship between these stimulus durations was reversed. Sue (1975) similarly found 5-second stimulus presentations to be superior to 30-second presentations with relaxation, but observed no effect of stimulus duration without relaxation. Testing his own theory, Watts (1971) found that, as predicted, anxiety reduction was greater for items low in phobic power if 5-second presentations were used rather than 30-second presentations; while for items high in phobic power this relation was reversed. This was the pattern of results on a short-term measure of anxiety reduction. But when long-term measures are taken, it seems generally to be the case that anxiety reduction is directly proportional to exposure duration (Mathews 1978), and this was observed in Watts' (1971) experiment even in the condition (low-intensity stimuli) in which, on the short-term measure, short durations produced a better effect. Watts (1979) suggests that this may be due to the more lasting effects of habituation relative to sensitization. Another finding consistent with the theory is that shortening the interval between stimulus presentations weakened the anxiety reduction observed with high-intensity stimuli, but had no effect with stimuli of low intensity (Watts 1973); this could be due to greater summation of sensitization over short inter-stimulus intervals when high-intensity stimuli are used (Watts 1979).

Clearly, this theory requires further confirmation. But it is well-enough specified to lend itself readily to experimental test; and it handles the existing data as well as any of its competitors. From the present point of view it has the great merit of fitting snugly with our theory of anxiety, both psychologically and physiologically. Thus, jointly, the two theories hold out the promise of a coherent conceptual framework extending from the anatomical basis of anxiety in the rat to its treatment by behavioural methods in man.

13.2 DRUG THERAPY

We commenced our consideration of treatment in the preceding section by focusing on the methods of behaviour therapy. A second important method of treating anxiety, of course, is with drugs (we took a brief look at the other major physical treatment, psychosurgery, at the start of Chapter 11). Since the cornerstone of the theory developed earlier in the book has been the assumption that drugs such as the benzodiazepines and the barbiturates or the later novel anxiolytics reduce anxiety, we cannot use the fact that they do (e.g. Rickels 1978) as evidence in favour of the theory. We did, however, make use of differences between the therapeutic profiles of drugs of different classes (e.g. the particular effectiveness of serotonin-uptake blockers in the treatment of panic disorder) to arrive at inferences concerning the brain sites most likely to mediate different anxiety-related symptoms and syndromes.

It is consistent with the evidence concerning the behavioural effects of the anxiolytic drugs in animals (Chapter 4) that they act only temporarily to suppress anxiety in man

(Lader and Marks 1972): they do not eliminate the patient's reaction to anxiogenic stimuli as, in the best of cases, does behaviour therapy. Indeed, it is even possible that, under some conditions, the anti-anxiety drugs antagonize the beneficial effects of behaviour therapy on the long-term reduction of anxiety. This might be expected to occur, given that barbiturates and benzodiazepines sometimes block the effects on resistance to extinction of both partial reinforcement and partial punishment (Feldon 1977; Gray 1977; Willner and Crowe 1977; Feldon *et al.* 1979; Davis *et al.* 1981; Feldon and Gray 1981). If these effects are construed as habituation to non-reward and punishment, respectively, and if behaviour therapy is construed as guided habituation to phobic stimuli, it follows that the anti-anxiety drugs may also block the effects of behaviour therapy.

The experimental literature relevant to this deduction is scarce and inconsistent (Marks 1976; Mathews 1978). Several findings, however, suggest that the deduction may be correct, at least for the classical anxiolytics.

Taub *et al.* (1977) report the results of an experiment with rats deliberately designed to mimic combinations of flooding and pharmacotherapy common in the clinic. They trained rats on a one-way avoidance task (jumping to a retractable ledge). There was then a single 10-minute session of exposure to the grid floor without shock and with the ledge retracted, equivalent in conception to a flooding session. Drugs were given only during this session. Three days later the rats were tested without shock to see to what extent their avoidance response had been affected by 'flooding'. There was a significant reduction in avoidance in undrugged, flooded rats relative to controls not given a flooding session. This effect (equivalent to anxiety reduction, if the model is valid) was attenuated in rats given chlordiazepoxide, amylobarbitone, or meprobamate, as would be predicted from the argument above. It is not certain, however, that this was due to a direct action of the drug, since state dependency may have played a role (no drug was administered before the test session). From a clinical point of view, of course, this theoretical nicety is unimportant, since the hope is to finish up with a patient free of both anxiety and drugs. In any case, both direct and state-dependent drug effects may contribute to the same end-result (weakened habituation), as in the case of the partial reinforcement extinction effect (Gray 1969; Feldon *et al.* 1979).

Clinical findings point in the same direction. Thus, Hafner and Marks (1976) treated agoraphobics by 'exposure *in vivo*', that is, by placing them in the real-life situation which elicited their anxiety. Exposure occurred after a placebo, or after diazepam given 1 or 4 hours previously (the 'peak' and 'waning' diazepam conditions, respectively). There was no effect of the drug if improvement was measured during or immediately after treatment. But at follow-up 6 months later the peak diazepam group was significantly worse than the placebo controls on measures of susceptibility to panic attacks and subjective anxiety, and non-significantly worse on measures of mood and somatic anxiety; and the waning diazepam group was not different from the placebo controls, but significantly superior to the peak diazepam group on three of these four measures. These results are in accord with expectation, both in that the deleterious effect of the drug was found in that condition (peak diazepam) which maximizes the likelihood of state dependency and in that it affected specifically long-term improvement. An earlier study by the same group (Marks *et al.* 1972) used essentially the same design with patients suffering from specific phobias (of blood, spiders, etc.), but there was no long-term follow-up. In this

experiment the immediate therapeutic effect of flooding *in vivo* was greater in the waning diazepam condition than in either of the other two. But this could have been due to differences in the amount of exposure, since the patients were tested individually and the speed of touching the phobic object was faster in the drugged subjects; as we have seen, increased exposure time would be expected to facilitate behaviour therapy. In the Hafner and Marks (1976) experiment, by contrast, patients were exposed in groups mixed with respect to drug condition, so this variable was controlled.

Thus there are probably two effects of the drug. One, beneficial, works by increasing the patient's exposure to the phobic object; the other, harmful, blocks the habituation that results from this exposure. In addition, of course, there is an acute reduction in anxiety while the drug continues to act directly. The interaction between these different effects may account for the mixed results obtained in other experiments of this kind (Hussain 1971; Johnston and Gath 1973; Chambless *et al.* 1979). But it is clear that anti-anxiety medication may sometimes attenuate or even reverse the benefits of behaviour therapy (Hafner and Marks 1976; Chambless *et al.* 1979). Given that the habituation which is specifically manipulated by behaviour therapy may occur in a less systematic manner also in the absence of therapy, so producing the high spontaneous recovery rate observed in anxiety syndromes, long-term maintenance on anti-anxiety drugs is usually contraindicated as a therapeutic strategy. The best use of these drugs is probably as a short-term crutch, either to aid coping with a particularly threatening situation or to facilitate exposure to phobic stimuli during behaviour therapy. But, at least in the present state of the art, simple drug therapy for anxiety should generally yield priority to behavioural methods of treatment.

These conclusions from earlier work are broadly supported by the results of two recent randomized, placebo-controlled clinical trials of the interactive effects of exposure therapy for agoraphobia and benzodiazepine administration, both with long-term follow-up. A strongly negative interaction (drug-induced antagonism of the gains due to behavioural treatment) has been reported when a high dose of alprazolam was administered (Marks *et al.* 1993); in contrast, when a low dose of diazepam was employed (Wardle *et al.* 1994), there were clinically significant effects of both exposure therapy and drug treatment, and the two effects did not interfere with each other (nor, however, were they additive). Clinically, therefore, it *is* possible to combine benzodiazepine treatment with behaviour therapy, although care must be taken, especially in the avoidance of high drug doses (although the possibility that the difference between the results of the Marks *et al.* and Wardle *et al.* trials was due to the specific benzodiazepine used, alprazolam as against diazepam, rather than dose, while unlikely, cannot be ruled out). Theoretically, the evidence from animal studies, reviewed above, that these compounds interfere in some manner with the processes underlying the development of behavioural tolerance for aversive events is (again with a *caveat* concerning dose) largely borne out by the results of the Marks *et al.* (1993) study. These results strongly caution, therefore, against long-term maintenance of patients on benzodiazepines since, as noted above, this may reduce the chances of spontaneous development of tolerance to adversity.

The blockade of the clinical effects of exposure therapy by alprazolam reported by Marks *et al.* (1993) provides support for the suggestion, made in our discussion of Watts' analysis of exposure therapy, that these effects are mediated by the septo-hippocampal

system. As noted above, the prediction that concomitant benzodiazepine treatment might disrupt the efficacy of exposure therapy was derived from the evidence in animal studies that these compounds block the effects of both partial reinforcement and partial punishment on resistance to extinction (Gray 1987b, Chapter 10). This line of argument leads to the inference that the neural loci at which anxiolytics block the partial reinforcement extinction effect are those at which, in the Marks *et al.* clinical study, alprazolam disrupted exposure therapy; and that therefore these loci participate in mediating the normal clinical effects of exposure therapy in undrugged patients. Animal studies clearly indicate the septo-hippocampal system and its ascending noradrenergic innervation as loci of this kind: destruction of the hippocampal formation, interruption of the output pathways linking the hippocampal formation to the nucleus accumbens, and destruction of the dorsal ascending noradrenergic bundle all block the partial reinforcement extinction effect; while electrical stimulation of the medial septal area so as to drive the hippocampal theta rhythm at frequencies selectively affected by anxiolytic drug administration is able to mimic this effect (see Appendices 7, 9, and 10 for reviews of the relevant data). At least in the case of agoraphobia (the condition studied by Marks *et al.* 1993), then, it seems likely that the septo-hippocampal system plays an important role in mediating the effects of exposure therapy.

13.3 COGNITIVE-BEHAVIOURAL THERAPY

The cognitive approach to the treatment of anxiety disorders is the newest arrival on the scene. Cognitive therapy is founded on a view of psychological disorder which postulates that people's cognitive appraisal of events can affect their behavioural and emotional responses to those events. In particular, maladaptive distortions in cognitive appraisals may lead to pathological emotions and behaviour (Ellis 1962; Beck 1976). Cognitive therapy aims to teach patients how to monitor and alter these maladaptive cognitions, and so to effect adaptive change in their emotions and behaviour. This type of treatment grew (starting in the 1970s) out of the same experimentally based approach to psychological science which had led earlier (in the 1950s) to the development of behaviour therapy. This aims, as we have seen above, to alter maladaptive behaviour directly, for example by exposing a phobic patient to anxiogenic stimuli; cognitive therapy, by contrast, attempts to alter behaviour by changing patterns of thought. For some years the two approaches were seen as competing. However, most clinicians (including certainly the majority of clinical psychologists in the United Kingdom) now employ a combination of specific behavioural and cognitive techniques, depending upon the particular problem requiring resolution. This approach is, therefore, termed 'cognitive-behavioural therapy'.

The most influential cognitive approach to the understanding and treatment of emotional disorders is that developed by A. T. Beck. Initially applied to depression, Beck's approach proposes that erroneous beliefs and dysfunctional information processing are not just a symptom of depression, but play a central role in its initiation and maintenance. Beck's (1970, 1976) cognitive therapy has been subjected to careful scrutiny and rigorous testing over the years and now has strong empirical support (Hawton *et al.* 1989;

Gloaguen *et al.* 1998). In his model 'cognitions' can take the form of either subvocal speech or images. They are defined as 'stream-of consciousness or automatic thoughts that tend to be in an individual's awareness' (Beck *et al.* 1983, p. 2). Automatic thoughts reflect the individual's appraisal of a situation (rather than the objective situation as such), and lead directly to emotional and behavioural responses. Distorted appraisals lead to dysfunctional responses. Distortions are particularly likely to arise when underlying dysfunctional 'schemata' have been activated. Schemata, finally, are 'stable, general, underlying beliefs and assumptions, which constitute a vulnerability to events' (Beck *et al.* 1983, p. 2). They are relatively unavailable to consciousness, and often laid down in childhood. Thus, automatic thoughts can be regarded as the surface manifestations of deeply held schemata.

Beck (1967, 1976) has described a number of specific types of cognitive distortions, such as dichotomous thinking (the tendency to think in absolutist, all-or-nothing terms), overgeneralization (unjustified extrapolation on the basis of a single instance), and arbitrary inference (jumping to conclusions when evidence is lacking or contradictory). These thinking errors, Beck suggests, are automatic, involuntary, and appear highly plausible to the thinker. In depression, they are manifest as the well-established negative cognitive triad: negative views of the self, current experience, and the future. Other cognitive changes maintain these views once they first occur. The negative cognitive shift, for example, influences perception, working memory, recall, and long-term memory (Beck 1987). The cognitive changes result in depressed mood and behaviour, and the lower mood in turn increases the probability of further negative automatic thoughts, so creating a vicious circle that maintains the depression.

Despite Beck's early emphasis on depression, his cognitive model represents a comprehensive formulation of psychopathology in general (Dobson and Block 1988). The model has been applied to a wide range of emotional disorders, including—of particular relevance here—anxiety (generalized anxiety disorder and panic), phobic disorders, and obsessions and compulsions, as well as, for example, eating disorders and drug abuse. Beck proposes that each of the various emotional disorders is causally related to systematic errors of thinking, and that these tend to be thematic; for example, depression is related to thoughts about loss, anxiety to themes of threat or danger, obsessive–compulsive disorder to beliefs about responsibility and guilt, and so on. There is no need to see these cognitive views of anxiety and obsessive–compulsive disorder as conflicting with the understanding of these conditions that we reached from a neural-system perspective in Chapter 11. Indeed, the postulated relationship between anxiety and distorted thinking about threat or danger is highly congruent with the general view of anxiety developed throughout this book (see, for example, the emphasis placed in Chapter 3 on cognitions rather than stimuli). More generally, cognitions can only function by being represented in trains of nervous impulses somewhere in the brain.

Cognitive-behavioural therapy is active and focused on the present (in contrast to psychodynamic psychotherapy, with its emphasis on early experience). The therapist designs specific learning experiences to teach clients how to monitor automatic thoughts; recognize the connections between cognition, emotion, and behaviour; collect evidence and generate alternative interpretations; substitute more realistic cognitions for distorted thoughts; and identify dysfunctional beliefs that predispose the individual to distorted interpretations. Generally, therapy is time-limited, relatively short-term in application,

but aiming at long-term changes in habits of thought that will contribute not only to recovery from the current psychiatric episode but also to prophylaxis against relapse. Fifteen to twenty 50-minute sessions at weekly intervals have been shown to be sufficient for both recovery from depression (Sacco and Beck 1985) and prophylactic effects (see below). Based upon the initial cognitive therapy for depression, a series of therapies has been developed for other, different emotional disorders. In each case, the general cognitive therapeutic strategy is the same: monitoring and altering dysfunctional thinking errors. In each case, also, behaviour therapy provides further specific tools; for example, graded exposure to feared stimuli for phobias and anxiety, or response prevention for obsessive–compulsive behaviour.

From their earliest beginnings both behaviour therapy and cognitive therapy have been strongly evidence-based, and this has continued to be true for their joint offspring, cognitive-behavioural therapy. Indeed, this type of treatment (taken here to include also its more purely cognitive forerunner) is now one of the most widely researched of all psychiatric treatments, psychotherapeutic or otherwise. Its efficacy, compared with other forms of both psychotherapy and pharmacotherapy, has been demonstrated repeatedly. Four major studies have shown it to be as effective as pharmacotherapy in the treatment of unipolar depression (Rush *et al.* 1977; Blackburn *et al.* 1981; Murphy *et al.* 1984; Evans *et al.* 1992), and more effective than pharmacotherapy in the prevention of relapse (Kovacs *et al.* 1981; Blackburn *et al.* 1986; Simons *et al.* 1986; Evans *et al.* 1992; Gloaguen *et al.* 1998). Having established its credentials in the difficult field of depression (where purely behavioural methods have been ineffective), cognitive-behavioural therapy was ready to roll out to other conditions. It has been shown to be effective also in the treatment of generalized anxiety disorder, for which several controlled trials have demonstrated its superiority over both pharmacotherapy (e.g. Power *et al.* 1990) and simple behaviour therapy (Butler *et al.* 1991). It is particularly effective in the treatment of panic disorder, both alone (in recent studies approximately 80% of panic disorder patients were panic-free after 3 months of treatment) and in combination with pharmacotherapy (Clark 1991). These effects persisted over 12 months follow-up. Social phobia, too, has recently succumbed to the therapeutic effects of this approach (Clark 1997), and other advances (e.g. in the treatment of eating disorders; Fairburn *et al.* 1991; Agras *et al.* 1992) are surely on their way.

It will be instructive to take a closer look at the way in which cognitive-behavioural therapy has been applied to panic disorder and social phobia. In Chapter 11 we related these two conditions to largely different neural substrates. Yet cognitive-behavioural therapy appears to work well and according to the same general principles in both cases. These common treatment effects reinforce the view we reached at the end of the previous chapter concerning the role played by personality: there are factors which operate on the defence system as a whole and which can, therefore, affect simultaneously all the different neural systems we were so careful to dissect from one another in Chapter 11. Furthermore, analysis of the processes that underlie cognitive therapy shows these to be highly congruent with the analysis of anxiety developed in this book.

Our discussion of the application of cognitive therapy to panic disorder and social phobia is based upon Clark's (1997) description of work on both conditions, much of it conducted in his own group in Oxford. We start, as indeed does the cognitive

approach to treatment itself, by outlining a cognitive model of the pathological processes thought to underlie each of these disorders.

The cognitive model of panic disorder states that:

> Individuals who experience recurrent panic attacks do so because they have a relatively enduring tendency to interpret certain bodily sensations in a catastrophic fashion. The sensations that are misinterpreted are mainly those involved in normal anxiety responses (e.g. palpitations, breathlessness, dizziness, paresthesias) but also include some other sensations. The catastrophic misinterpretation involves perceiving these sensations as much more dangerous than they really are and, in particular, interpreting the sensations as indicative of *immediately* impending physical or mental disaster—for example, perceiving a slight feeling of breathlessness as evidence of impending cessation of breathing and consequent death, perceiving palpitations as evidence of an impending heart attack, perceiving a pulsing sensation in the forehead as evidence of brain haemorrhage, or perceiving a shaky feeling as evidence of impending loss of control and insanity. (Clark 1988, p. 149.)

This sequence of events is illustrated in Fig. 13.1.

> External stimuli (such as a department store for an agoraphobic) and internal stimuli (body sensations, thoughts, images) can both provoke panic attacks. The sequence that culminates in an attack starts with the stimuli being interpreted as a sign of impending danger. This interpretation produces a state of apprehension, which is associated with a wide range of bodily sensations. If these anxiety-produced sensations are interpreted in a catastrophic fashion (impending insanity, death, loss of control, etc.) a further increase in apprehension occurs, producing more bodily sensations, leading to a vicious circle which culminates in a panic attack. (Clark 1997, pp. 124–5.)

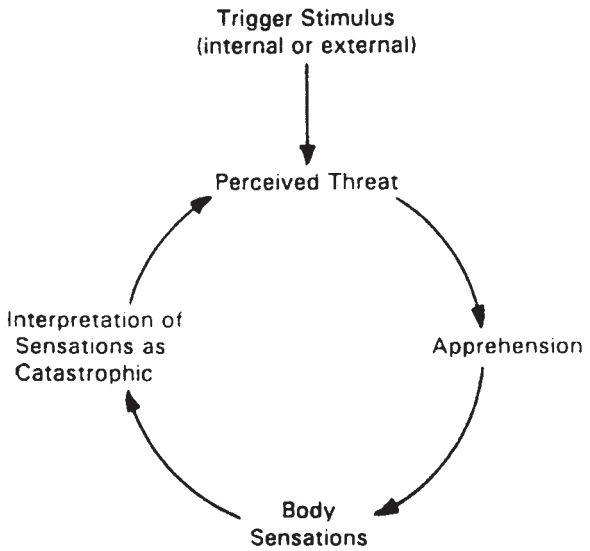


Fig. 13.1 The suggested vicious circle underlying a panic attack. (From Clark 1986.)

Further consequences of these tendencies to interpret bodily sensations in a potentially catastrophic fashion are that the subject becomes hypervigilant and repeatedly scans his inner sensations for signs of threat; and that he develops various patterns of 'safety' behaviour that avoid exposure to these threatening signs, but at the same time preclude any kind of disconfirmation of the erroneous beliefs that led to their being threatening in the first place. So, a patient afraid that he is suffering from heart disease will avoid exercise, precluding discovery of the fact that his heart rate may rise without the feared heart attack ensuing.

A central feature of this cognitive model of panic is, then, that attacks are due to an excessive weighting of adverse outcomes possibly associated with the stimuli that trigger the trains of thought illustrated in Clark's vignettes. This analysis of the way in which panic attacks arise has been substantiated by much empirical research, reviewed by Clark (1996, 1997), as well as by the success of the therapy based upon it.

The cognitive model of social phobia (Clark and Wells 1995; Clark 1997; see Fig. 13.2) shares many features with the cognitive model of panic.

> Social phobics are said to develop a series of assumptions about themselves and social situations. For example: 'Unless someone shows that they like me, they dislike me. Unless I am liked by everyone, I am worthless. If I show I am anxious people will think I am odd/will reject me.' These assumptions lead them to interpret normal social interactions in a negative way, viewing them as signs of danger. For example, if a social phobic is talking to someone at a party and the other person briefly looks out of the window the social phobic may think 'I'm being boring'. This interpretation triggers an 'anxiety programme' that can be usefully divided into three interlinked components. (Clark 1997, pp. 126–7.)

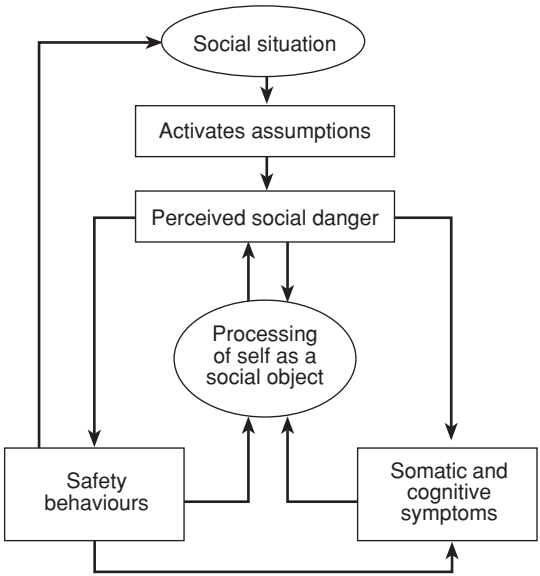


Fig. 13.2 The processes hypothesized to occur when a social phobic enters a feared situation. (From Clark and Wells 1995.)

These three components are described by Clark (1997) as consisting of: 'the somatic and cognitive symptoms of anxiety that are reflexively triggered by the perception of danger' (e.g. blushing, trembling, racing heart, etc.); 'safety behaviours which patients engage in to try to reduce social threat and prevent feared outcomes from occurring' (e.g. trying not to attract attention, avoiding eye contact, censoring what one says, etc.); and a shift in attention, i.e. 'when social phobics think they are in danger of negative evaluation by others, they shift their attention to detailed monitoring and observation of themselves'. These are the same three components that operate in the cognitive model of panic (hypervigilance to bodily sensations in panic disorder being replaced, for example, by hypervigilance to aspects of one's own social performance in social phobia, i.e. in each case to the relevant set of threat stimuli). In this way, a vicious circle is again set in motion, since *good* social performance is impeded by the very symptoms of anxiety, safety behaviours, and self-observation to which the social phobic is driven by fear of *poor* social performance. And, as before, we may see the whole process as depending upon the excessive weighting given to the potential adverse outcomes associated with specific stimuli, those stemming from social interaction taking, in social phobia, the place of bodily sensations in panic disorder. (For the empirical evidence supporting this cognitive model of social phobia, the reader is again referred to Clark 1997.)

We have indicated above the general ways in which cognitive-behavioural therapy re-educates the patient into altering maladaptive belief systems, with consequent desired change in both emotional and behavioural responses. Application of these general principles to panic disorder and social phobia is reasonably straightforward (although the development of effective therapies has required in practice a detailed and painstaking putting together of many specific procedures, often insightful to the point of inspiration; this too is reviewed in Clark 1996, 1997). For panic disorder, the key steps include: identifying the specific triggers for panic in the particular patient; the collation of evidence contradicting the patient's maladaptive beliefs about, for example, the likelihood of his having a heart attack; behavioural experiments, for example, inducing feared bodily sensations in the clinic and demonstrating that no adverse consequences follow, or preventing safety behaviours to allow patients to disconfirm their expectations of adverse outcomes. *Mutatis mutandis*, similar steps are involved in the treatment of social phobia. For example, the patient may be asked to play a social role with and without the aid of his usual repertoire of safety behaviours while being videotaped, so that there is a record for later discussion with the therapist. The success rates with these therapies have averaged, for panic disorder, 80% panic-free over a total of 97 patients across five studies (see Table 6.1 in Clark 1997); while the more recent application to social phobia has so far achieved rates of improvement up to 65% (Heimberg *et al.* 1990; and see Clark 1997, p. 148). In the present context, the key feature of cognitive-behavioural therapy is that it attempts to reduce (in the ways outlined above) the excessive weighting of potential adverse outcomes that gives rise to panic disorder and social phobia in the first place. In this way, cognitive-behavioural therapy attacks these disorders at the root, in contrast to pharmacotherapy, which deals only with symptoms. The effects of cognitive-behavioural therapy, furthermore, are long-lasting, again in contrast to pharmacotherapy, whose beneficial effects are usually lost when treatment terminates.

Where in the brain might cognitive-behavioural therapy act to prevent excessive weighting from being given to potential adverse outcomes? It is tempting to attribute this effect to a reduction in the activity of the septo-hippocampal system. This conclusion would be congruent with the general manner in which, in this book, we have seen the septo-hippocampal system as working to *increase* the power of adverse associations. Cognitive-behavioural therapy could then be understood, neatly, as directly reversing the fundamental pathological process underlying anxiety. This conclusion would agree, also, with the similar one we drew above in relation to the probable neural substrate for the effectiveness of exposure therapy in the treatment of agoraphobia.

Unfortunately for neatness, however, there are two reasons why we cannot accept this conclusion. The first is that the effectiveness of cognitive-behavioural therapy is too broad. It is effective not only in conditions, such as generalized anxiety disorder, that our model sees as centrally involving anxiety and the activity of the septo-hippocampal system, but also in those, such as panic disorder, which depend upon activity in other regions of the defence system and are associated with other emotional states. To put the same point differently: if cognitive-behavioural therapy improves panic disorder by reducing activity in the septo-hippocampal system, then our model commits us to the (erroneous) prediction that panic (not to mention depression, the condition in which this form of therapy cut its teeth) should respond to treatment with anxiolytic drugs (since we see these compounds too as working by reducing activity in the septo-hippocampal system). The second reason not to attribute the effectiveness of cognitive-behavioural therapy to action in the septo-hippocampal system is that the type of rational discourse that is central to cognitive therapy can be mediated only by way of the linguistic systems of the neocortex. These are the same systems to which we have attributed a descending control (by way of the prefrontal and cingulate cortices) over the septo-hippocampal system, accounting for the existence of forms of anxiety that are resistant to the therapeutic effects of the anxiolytic drugs, but which benefit from the type of psychosurgery that severs these descending pathways (see Chapter 11). One might, therefore, suppose that, even though the septo-hippocampal system is unlikely directly to mediate the effects of cognitive-behavioural therapy, it nonetheless lies on the final common pathway that links these neocortical systems to, for example, the central grey, providing cognitive control, both facilitatory and inhibitory, over panic attacks. Unfortunately, however, this hypothesis continues to predict clinical efficacy in panic disorder for the anxiolytics (acting on this final common pathway); and no such efficacy is observed.

The fact that cognitive-behavioural therapy is clinically effective in at least three disparate anxiety-related disorders—generalized anxiety disorder (Power *et al.* 1990), panic disorder, and social phobia (Clark 1997)—suggests that these conditions share a common core upon which this form of therapy can act, despite their different symptoms. We have just rejected the possibility that this common core might consist in a *reduction* in the activity of the septo-hippocampal system. But there is an interesting inversion of this hypothesis that deserves consideration.

The arguments we pursued in Chapter 11 stressed the particular brain regions, activity in which proximately underlies each disorder-specific symptom—the central grey for symptoms of panic, the amygdala for the detection of negative cues of social interaction, and the septo-hippocampal system itself for the diffuse and pervasive worry of generalized

anxiety disorder. The neurology we outlined in Chapter 11, however, had another important feature: the strongly recursive interactions that link the different structures making up both the primary defence system (central grey, hypothalamus, amygdala) and the anxiety system (septo-hippocampal system, amygdala, cingulate and prefrontal cortices). As argued in that chapter, the septo-hippocampal system occupies a nodal position in these recursive interactions. In particular, it appears to exercise both excitatory and inhibitory descending influences over the primary defence system. The descending excitatory influences lead to the widespread array of physiological and behavioural symptoms that can arise from even highly abstract sources of worry in, for example, obsessive–compulsive disorder or generalized anxiety disorder; the descending inhibitory influences are able to restrain panic behaviour (as analysed, for example, in Graeff's work; see Section 11.12).

Cognitive-behavioural therapy perhaps recruits this latter route to achieve its beneficial effects. If so, we would not expect this form of treatment to act by reducing activity in the septo-hippocampal system, but rather by redirecting and in some cases even enhancing it. So, in generalized anxiety disorder or social phobia, the effectiveness of cognitive-behavioural therapy perhaps depends upon altered inputs to the septo-hippocampal system consequent upon neocortically mediated, verbally coded redescriptions of relevant stimuli, leading to altered outputs from the system, for example, to the amygdala. In panic disorder, the altered outputs from the septo-hippocampal system could take the form of increased inhibitory control over lower centres mediating panic attacks. Note that this 'inverted' form of the initial hypothesis that the septo-hippocampal system mediates, at least in part, cognitive-behavioural therapy no longer leads to the erroneous prediction that anxiolytic drugs should be effective in panic disorder. Rather, it predicts that, if classical anxiolytics were administered concomitantly with cognitive-behavioural therapy, the clinical effects of the latter (like those of behaviour therapy; see above) would be impeded. There are no negative interactions, and some evidence of a modest advantage, when cognitive-behavioural therapy is combined with anti-depressant medication (Hollon *et al.* 1991, 1992); but we know of no data relative to its combination with the classical anxiolytics. A further prediction the inverted hypothesis makes is that the effectiveness of cognitive-behavioural therapy should be weakened in patients with panic disorder or post-traumatic stress disorder who have suffered damage to the hippocampus (see Sections 11.12 and 11.14); this hypothesis can be readily tested using contemporary techniques of structural neuroimaging. But there is, in addition, a range of alternative (and not necessarily mutually exclusive) descending routes from the frontal and cingulate cortices to, for example, the central grey (see Appendix 2), a particularly plausible route when cognitive-behavioural therapy is applied to panic disorder; or to the serotonergic dorsal raphe nucleus and noradrenergic locus coeruleus (see Appendix 10), with consequent re-ascending influences that have widespread effects throughout the entire forebrain.

13.4 CODA

In the stock phrase, there has been an explosion of new data in the fields covered in this book since its first edition in 1982. At times we have feared that the theory advanced in that first edition would explode along with the new data. It has certainly suffered

considerable strain. In 1982 we were able to concentrate more or less exclusively on a single, unified behavioural inhibition system, with an almost equally unified neural system mediating its functions and thus underlying the occurrence of anxiety. That unity has come under attack from several directions.

First, as set out in Chapter 1, came the recognition that, alongside our preferred starting point for the construction of a neuropsychology of anxiety (namely, the behavioural effects of the anti-anxiety drugs), we needed to take due account of the implications emerging from three alternative starting points: (1) the fight–flight behaviour mediated by the central grey and hypothalamus; (2) conditioned fear, mediated by the amygdala; and (3) the roles played by the cingulate and frontal cortices in drug-resistant human anxiety. An initial resolution of the tension created by the first of these alternatives, building on Klein's (1981) clinical observations and Graeff's (1994) experimental studies, was to distinguish categorically between panic and anxiety, and to allocate the mediation of the former to the central grey and hypothalamus and the latter to the septo-hippocampal system (Gray 1987b). Somewhat similarly, in partial resolution of the tension created by the third alternative, we built on Rapoport's (1989) work on obsessive–compulsive disorder so as to ascribe to the cingulate cortex a role in this condition analogous to that played by the basal ganglia in disorders of motor function. Also, driven in part by our own psychopharmacological observations (Sartory *et al.* 1990) as well as the accumulating evidence (LeDoux 1996) for the crucial role played by the amygdala in fear, we ascribed specific phobias to this structure. But the price paid for these resolutions, in terms of theoretical fragmentation, was not negligible. Where, in 1982, we dealt with one clinical state (that of anxiety), distinguishing only between elicitation of this state upstream (by way of the ascending monoaminergic pathways, and drug-sensitive) or downstream (by way of descending projections from linguistically competent regions of the neocortex, and drug-resistant), we now have additional 'anxiety-related' disorders—panic, specific phobia, and obsessive–compulsive behaviour (and others; see Chapter 11)—that are not the same as anxiety itself.

Second, even within the conceptual framework provided by the behavioural inhibition system, there is additional theoretical fragmentation. In part, this comes from the recognition that some of the functions of this system (especially its increased arousal and autonomic outputs) need to be allocated to the amygdala (not seen in 1982 as forming part of this system). This in itself would not pose a major problem, since the 1982 treatment of the behavioural inhibition system fully accepted that its neural basis was distributed and extended more widely than the septo-hippocampal system (with this in turn having functions wider than an involvement exclusively in anxiety). Indeed, joining together the septo-hippocampal system and amygdala in this way to provide an account of the neural basis of anxiety has several advantages (not least, that it links our theory with the important work on the involvement of the amygdala in emotion carried out in LeDoux's and Davis's laboratories). In particular, it integrates the 'feeling' (amygdalar) and cognitive (septo-hippocampal) aspects of anxiety in a manner that is clearly required by the clinical data, thus avoiding the sterile debates that divide those who seek to emphasize one of these aspects at the expense of the other.

However, as discussed at various points above (see especially Section 11.3), recruitment of the amygdala into the behavioural inhibition system causes other problems.

Whereas in 1982 we could see a reasonably simple way in which, by acting on ascending monoaminergic (noradrenergic and serotonergic) pathways to alter septal control of the hippocampal theta rhythm, anxiolytic drugs could act in a unified manner upon the whole of the distributed neural system, addition of the amygdala (together with increased understanding of the mechanisms by which anxiolytics cause changes in the hippocampal theta rhythm) renders this theoretical stratagem impossible. Thus, to account now for the behavioural effects of the anti-anxiety drugs, we need to allot them apparently independent modes of action additional to that on the ascending monoaminergic pathways: on the supramammillary nucleus, altering the frequency setting of theta, and on the amygdala, altering such phenomena as fear-potentiated startle.

Given these developments, why should we persist in seeing *any* form of unity in the neural basis of anxiety and the anxiety-related disorders? Perhaps the whole notion of anxiety is a spurious way of classifying together a hopelessly disparate set of phenomena, nothing more than a heading in a textbook. The data on predisposition to the comorbidity of the anxiety-related disorders, on the heritability of this personality trait, and on the sensitivity of anxiety-related disorders to cognitive-behavioural treatment, reviewed in this and the previous chapter, offer apparent antidotes to this counsel of despair. Strikingly, the unity suggested by these sources of data binds both the states that we have considered as the core components of anxiety, whether normal or pathological (as, for example, in generalized anxiety disorder), and those we have seen as lying outside anxiety proper, such as panic disorder. What appears to be inherited is a set of enduring personality traits (measured as the degree of Neuroticism) that predisposes to any of these conditions; and all appear to a significant degree to be responsive to cognitive-behavioural therapy.

It would be comforting if we could infer from these data that, if we just work a little harder and longer, we shall find in the neuropsychology of anxiety a common core that is genetically determined, neurologically unified, and manipulated by cognitive-behavioural therapy. However, this inference is probably illegitimate, and the legitimate one less interesting. The following somatic analogy may help to make this point.

An inherited low level of immune system function could predispose people to catch a variety of bacterial infections. These infections are extrinsic rather than intrinsic, are not the direct result of the inherited bias, and will differ markedly in their symptomatology, pathological characteristics, and required nursing. However, all might be treatable by the same or related antibiotics (for reasons that are independent of the nature of the inherited low level of immune function). This analogy shows how there can be a clear link between common predisposing factors and common treatments which relate to common properties of all the disease organisms, but which nonetheless allows the different diseases to be caused by different organisms and to involve different pathological pathways in different bodily organs. Pursuing the analogy further, such an inherited 'predisposition to infection' is not necessarily maladaptive (as we have argued anxiety also not to be), since it could result in a low level of autoimmune disease.

Bearing this analogy in mind, then, what might be inherited as the predisposing factor of Neuroticism; and what might be the common mode of action of cognitive-behavioural therapy in all the anxiety-related disorders? One possibility, touched upon in the previous chapter, is that Neuroticism reflects principally a *perceptual* bias (that is to say, a *cognitive*

bias since, as emphasized in this chapter, perception requires interpretation of what is perceived) towards the identification or magnification of threat of all kinds. As emphasized in LeDoux's hierarchical system (see Fig. 1.8), threat is received at all levels of the defensive system (those which then respond most strongly depending upon, for example, defensive distance, as described in Chapter 2). The nature of the behaviour displayed, whether normal or requiring psychiatric treatment, and the specific emotion felt will depend in addition upon such factors as defensive direction, the presence of conflict, etc. as considered *in extenso* in this book. The degree of bias towards the perception of threat can then be seen as being influenced by polygenes, as discussed in the previous chapter, as well as by conditioning or trauma (as, for example, in the long-lasting changes observed in post-traumatic stress disorder, discussed in Chapter 11). With regard to psychological treatment, we can see this—as set out in this chapter—as acting (either by habituation in exposure therapy or by cognitive restructuring in cognitive therapy) directly to reduce such excessive perception of threat, so affecting both fear and anxiety. Anxiolytic drug treatment, in contrast, would alter the increased negative biasing associated with conflict, and so reduce anxiety only.

In some such way as this, then, the unity binding the anxiety-related disorders that is suggested by the data on personality, heritability, and psychological treatments can be reconciled with the diversity of underlying neural mechanisms that has occupied our attention for most of this book. That diversity remains (as does the enigma posed by the apparently selective and unified action of the anxiolytic drugs)—it has not been spirited away. However, if the suggestions made in the previous chapter are correct, then polygenes (or some of them) that determine the genetic contribution to Neuroticism will turn out to amplify the operation of systems that detect threat generally, and thus the entirety of the networks that subserve defence. And, if the suggestions made in the present chapter are also correct, then cognitive-behavioural therapy works in the reverse manner, by dampening the operation of these same systems.

That is as far as we can at present take the argument. The coming years will surely be no less explosive than those since 1982. We wait with interest to see whether they will blow our theory away.

References

Abelson, J. L. and Nesse, R. M. (1994). Pentagastrin infusions in patients with panic disorder. I. Symptoms and cardiovascular responses. *Biological Psychiatry*, **36**, 73–83.

Aboitiz, F. (1995). Working memory networks and the origin of language areas in the human brain. *Medical Hypotheses*, **44**, 504–6.

Abraham, W. C. (1988). Long-term potentiation as a possible associative memory mechanism in the brain. *New Zealand Journal of Psychology*, **17**, 49–58.

Abraham, W. C. and Bear, M. F. (1996). Metaplasticity: the plasticity of synaptic plasticity. *Trends in Neuroscience*, **19**, 126–30.

Ackil, J. E., Mellgren, R. L., Halgren, C., and Frommer, G. P. (1969). Effects of CS pre-exposures on avoidance learning in rats with hippocampal lesions. *Journal of Comparative and Physiological Psychology*, **69**, 739–47.

Adamec, R. E. (1998a). Evidence that NMDA-dependent limbic neural plasticity in the right hemisphere mediates pharmacological stressor (FG-7142)-induced lasting increases in anxiety-like behavior. Study 1. Role of NMDA receptors in efferent transmission from the cat amygdala. *Journal of Psychopharmacology*, **12**, 122–8.

Adamec, R. E. (1998b). Evidence that NMDA-dependent limbic neural plasticity in the right hemisphere mediates pharmacological stressor (FG-7142)-induced lasting increases in anxiety-like behavior. Study 2. The effects on behavior of block of NMDA receptors prior to injection of FG-7142. *Journal of Psychopharmacology*, **12**, 129–36.

Adamec, R. E. (1998c). Evidence that NMDA-dependent limbic neural plasticity in the right hemisphere mediates pharmacological stressor (FG-7142)-induced lasting increases in anxiety-like behavior. Study 3. The effects on amygdala efferent physiology of block of NMDA receptors prior to injection of FG-7142 and its relationship to behavioral change. *Journal of Psychopharmacology*, **12**, 227–38.

Adolphs, R., Tranel, D., Damasio, H., and Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, **372**, 669–72.

Aggleton, J. P. (1992). *The amygdala*. Wiley-Liss, New York.

Aggleton, J. P. (1993). The contribution of the amygdala to normal and abnormal emotional states. *Trends in Neuroscience*, **16**, 328–33.

Aggleton, J. P., Hunt, P. R., and Shaw, C. (1990). The effects of mammillary body and combined amygdalar–fornix lesions on tests of delayed non-matching-to-sample in the rat. *Behavioural Brain Research*, **40**, 145–57.

Agras, W. S. (1985). *Panic: facing fears phobias and anxiety*. W. H. Freeman, New York.

Agras, W. S., Rossiter, E. M., Arnow, B., Schneider, J. H., Telch, C. S., Raeburn, S. D., *et al.* (1992). Pharmacologic and cognitive-behavioural treatment for bulimia nervosa: a controlled comparison. *American Journal of Psychiatry*, **149**, 82–7.

Akhtar, S., Wig, N., Varma, O., Pershad, D., and Verma, S. (1975). A phenomenological analysis of symptoms in obsessive–compulsive neurosis. *British Journal of Psychiatry*, **127**, 342–8.

Akil, H., Watson, S. J., Young, E., Lewis, M. E., Khachaturian, H., and Walker, J. M. (1984). Endogenous opioids: biology and function. *Annual Review of Neuroscience*, **7**, 223–55.

Alemayehu, S., Bergey, G. K., Barry, E., Krumholz, A., Wolf, A., Fleming, C. P., *et al.* (1995). Panic attacks as ictal manifestations of parietal lobe seizures. *Epilepsia*, **36**, 824–30.

Alvarado, M. C. and Rudy, J. W. (1995a). Rats with damage to the hippocampal-formation are impaired on the transverse-patterning problem but not on elemental discriminations. *Behavioral Neuroscience*, **109**, 204–11.

Alvarado, M. C. and Rudy, J. W. (1995b). Comparison of 'configural' discrimination problems: implications for understanding the role of the hippocampal formation in learning and memory. *Psychobiology*, **23**, 178–84.

Alvarez, P., Zola-Morgan, S., and Squire, L. R. (1994). The animal model of human amnesia: long-term memory impaired and short-term memory intact. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 5637–41.

Alvarez, P., Zola-Morgan, S., and Squire, L. R. (1995). Damage limited to the hippocampal region produces long-lasting memory impairment in monkeys. *The Journal of Neuroscience*, **15**, 3796–807.

Amaral, D. G., Price, J. L., Pitkänen, A., and Carmichael, S. T. (1992). Anatomical organization of the primate amygdaloid complex. In *The amygdala: neurobiological aspects of emotion, memory and mental dysfunction* (ed. J. P. Aggleton), pp. 1–66. Wiley-Liss.

American Psychiatric Association (1987). *DSM-III-R: Diagnostic and statistical manual of mental disorders* (3rd edn, revised).

American Psychiatric Association (1994). *DSM-IV: Diagnostic and statistical manual of mental disorders* (4th edn).

Amsel, A. (1962). Frustrative nonreward in partial reinforcement and discrimination learning: some recent history and a theoretical extension. *Psychological Review*, **69**, 306–28.

Amsel, A. (1972). Behavioural habituation, counterconditioning, and a general theory of persistence. In *Classical conditioning II: current research and theory* (ed. A. H. Black and W. F. Prokasy), pp. 409–26. Appleton-Century-Crofts, New York.

Amsel, A. (1992). *Frustration theory: an analysis of dispositional learning and memory*. Cambridge University Press.

Andersen, P., Bliss, T. V. P., and Skrede, K. K. (1971). Lamellar organization of hippocampal excitatory pathways. *Experimental Brain Research*, **13**, 222–38.

Anderson, J., Huppert, F., and Rose, G. (1993). Normality, deviance and minor psychiatric morbidity in the community. *Psychological Medicine*, **23**, 475–85.

Andrews, G., Stewart, G., Morris-Yates, A., Holt, P., and Henderson, S. (1990). Evidence for a general neurotic syndrome. *British Journal of Psychiatry*, **15**, 6–12.

Andrews, G., Crino, R., Hunt, C., Lampe, L., and Page, A. (1994). *The treatment of anxiety disorders*. Cambridge University Press.

Antony, M. M., Brown, T. A., and Barlow, D. H. (1997). Response to hyperventilation and 5.5% CO_2 inhalation of subjects with types of specific phobia, panic disorder, or no mental disorder. *American Journal of Psychiatry*, **154**, 1089–95.

Aoki, Y., Fujihara, S., and Kitamura, T. (1994). Panic attacks and panic disorder in a Japanese non-patient population: epidemiology and psychosocial correlates. *Journal of Affective Disorders*, **32**, 51–9.

Arnold, M. B. (1968). *The nature of emotion*. Penguin Books, Harmondsworth.

Ashton, H. (1994). Guidelines for the rational use of benzodiazepines: when and what to use. *Drugs*, **48**, 25–40.

Asmundson, G. J. G. and Stein, M. B. (1994). Triggering the false suffocation alarm in panic disorder patients by using a voluntary breath-holding procedure. *American Journal of Psychiatry*, **151**, 264–6.

Baddeley, A. (1981). Amnesia: a minimal model and an interpretation. In *Human memory and amnesia* (ed. L. Cermak), pp. 305–36. Erlbaum, Hillsdale, NJ.

Baddeley, A. (1986). *Working memory*. Clarendon Press, Oxford.

Baddeley, A. (1988). Cognitive psychology and human memory. *Trends in Neuroscience*, **11**, 176–81.

Baddeley, A. and Wilson, B. A. (1994). When implicit learning fails: amnesia and the problem of error elimination. *Neuropsychologia*, **32**, 53–68.

Baenninger, R. (1967). Contrasting effects of fear and pain on mouse killing by rats. *Journal of Comparative and Physiological Psychology*, **63**, 298–303.

Baer, L. and Minichiello, W. E. (1990). Behavioural treatment (BT) for obsessive–compulsive disorder. In *Handbook of anxiety*, Vol. 4, *The treatment of anxiety* (ed. R. Noyes, Jr., M. Roth, and D. Burrows), pp. 363–87. Elsevier.

Baker, D. G., West, S. A., Orth, D. N., Hill, K. K., Nicholson, W. E., Ekhator, N. N., *et al.* (1997). Cerebrospinal fluid and plasma beta-endorphin in combat veterans with post-traumatic stress disorder. *Psychoneuroendocrinology*, **22**, 517–29.

Balaban, M. T. and Taussig, H. N. (1994). Salience of fear/threat in the affective modulation of the human startle blink. *Biological Psychology*, **38**, 117–31.

Ball, G. G. and Gray, J. A. (1971). Septal self-stimulation and hippocampal activity. *Physiology and Behavior*, **6**, 547–9.

Bandler, R. and Carrive, P. (1988). Integrated defence reaction elicited by excitatory amino acid microinjection in the midbrain periaqueductal grey region of the unrestrained cat. *Brain Research*, **439**, 95–106.

Barbas, H. (1995). Pattern in the cortical distribution of prefrontally directed neurons with divergent axons in the rhesus monkey. *Cerebral Cortex*, **2**, 158–65.

Barbas, H. and Pandya, D. N. (1987). Architecture and frontal cortical connections of the premotor cortex (area 6) in the Rhesus monkey. *The Journal of Comparative Neurology*, **256**, 211–28.

Barbas, H. and Pandya, D. N. (1991). Patterns of connections of the prefrontal cortex in the rhesus monkey associated with cortical architecture. In *Frontal lobe function and dysfunction* (ed. H. S. Levin, H. M. Eisenberg, and A. L. Benton). Oxford University Press.

Barker, J. L. and Ransom, B. R. (1978). Pentobarbitone pharmacology of mammalian central neurones grown in tissue culture. *Journal of Physiology (London)*, **280**, 355–72.

Barnes, C. A. (1988). Spatial learning and memory processes: the search for their neurobiological mechanisms in the rat. *Trends in Neuroscience*, **11**, 163–9.

Barnes, C. A., McNaughton, B. L., Mizumori, S. J. Y., Leonard, B. W., and Lin, L.-H. (1990). Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing. *Progress in Brain Research*, **83**, 287–300.

Barnes, J. M., Barnes, N. M., and Cooper, S. J. (1992). Behavioural pharmacology of $5HT_3$-receptor ligands. *Neuroscience and Biobehavioral Reviews*, **16**, 107–13.

Barrett, J. E. and Vanover, K. E. (1993). 5HT receptors as targets for the development of novel anxiolytic drugs: models, mechanisms and future directions. *Psychopharmacology (Berlin)*, **112**, 1–12.

Barsky, A. J., Cleary, P. D., Sarnie, M. K., and Ruskin, J. N. (1994). Panic disorder, palpitations, and the awareness of cardiac activity. *Journal of Nervous and Mental Diseases*, **182**, 63–71.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., and Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science*, **269**, 1115–18.

Beck, A. T. (1967). *Depression: clinical, experimental and theoretical aspects*. Harper and Row, New York.

Beck, A. T. (1970). Cognitive therapy: nature and relation to behaviour therapy. *Behavior Therapy*, **1**, 184–200.

Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*. International Universities Press, New York.

Beck, A. T. (1987). Cognitive models of depression. *Journal of Cognitive Psychotherapy*, **1**, 5–37.

Beck, A. T., Epstein, N., and Harrison, R. (1983). Cognitions, attitudes, and personality dimensions in depression. *British Journal of Cognitive Psychology*, **1**, 1–16.

Beck, C. H. M. and Fibiger, H. C. (1995). Conditioned fear-induced changes in behavior and in the expression of the immediate early gene c-fos: with and without diazepam pretreatment. *The Journal of Neuroscience*, **15**, 709–20.

Beech, R. (1974). *Obsessional states*. Methuen, London.

Bennett, A. T. D. (1996). Do animals have cognitive maps? *The Journal of Experimental Biology*, **199**, 219–24.

Benton, A. L. (1991). The prefrontal region: its early history. In *Frontal lobe function and dysfunction*, (ed. H. S. Levin, H. M. Eisenberg, and A. L. Benton), pp. 3–32. Oxford University Press.

Berger, T. W. and Alger, B. (1976). Neuronal substrates of classical conditioning in the hippocampus. *Science*, **192**, 483–5.

Berger, T. W. and Thompson, R. F. (1978a). Neuronal plasticity in the limbic system during classical conditioning of the rabbit nictating membrane response. I. The hippocampus. *Brain Research*, **145**, 323–46.

Berger, T. W. and Thompson, R. F. (1978b). Neuronal plasticity in the limbic system during classical conditioning of the rabbit nictitating membrane response. II. Septum and mammillary bodies. *Brain Research*, **156**, 293–314.

Berger, T. W., Rinaldi, P. C., Weisz, D. J., and Thompson, R. F. (1983). Single-unit analysis of different hippocampal cell types during classical conditioning of rabbit nictitating membrane response. *Journal of Neurophysiology*, **50**, 1197–219.

Bermant, G. (1961a). Response latencies of female rats during sexual intercourse. *Science*, **133**, 1771–3.

Bermant, G. (1961b). Regulation of sexual contact by female rats. Unpublished thesis, Harvard University, cited by Carlson (1980).

Bersh, P. J. (1980). Eysenck's theory of incubation: a critical analysis. *Behaviour Research and Therapy*, **18**, 11–17.

Bignami, G. (1965). Selection for high rates and low rates of avoidance conditioning in the rat. *Animal Behaviour*, **13**, 221–7.

Bitran, D., Shiekh, M., Dowd, J. A., Dugan, M. M., and Renda, P. (1998). Corticosterone is permissive to the anxiolytic effect that results from the blockade of hippocampal mineralocorticoid receptors. *Pharmacology, Biochemistry and Behavior*, **60**, 879–87.

Blackburn, I. M., Bishop, S., Glen, A. I. M., Whalley, G. T., and Christie, P. J. (1981). The efficacy of cognitive therapy and pharmacotherapy, each alone and in combination. *British Journal of Psychiatry*, **139**, 181–9.

Blackburn, I. M., Eunson, K. M., and Bishop, S. (1986). A two-year naturalistic follow-up of depressed patients treated with cognitive therapy, pharmacotherapy and a combination of both. *Journal of Affective Disorders*, **10**, 67–75.

Blackman, D. E. (1983). On cognitive theories of animal learning: extrapolation from humans to animals. In *Animal models of human behaviour* (ed. G. C. L. Davey), pp. 37–50. Wiley, New York.

Blair, H. T. and Sharp, P. E. (1995). Anticipatory head direction signals in anterior thalamus: evidence for a thalamocortical circuit that integrates angular head motion to compute head direction. *The Journal of Neuroscience*, **15**, 6260–70.

Blanchard, D. C. and Blanchard, R. J. (1990). Effects of ethanol, benzodiazepines and serotonin compounds on ethopharmacological models of anxiety. In *Anxiety* (ed. N. McNaughton and G. Andrews), pp. 188–99. Otago University Press, Dunedin.

Blanchard, D. C., Williams, G., Lee, E. M. C., and Blanchard, R. J. (1981). Taming of wild *Rattus norvegicus* by lesions of the mesencephalic central gray. *Physiological Psychology*, **9**, 157–63.

Blanchard, R. J. and Blanchard, D. C. (1989). Antipredator defensive behaviors in a visible burrow system. *Journal of Comparative Psychology*, **103**, 70–82.

Blanchard, R. J. and Blanchard, D. C. (1990a). An ethoexperimental analysis of defense, fear and anxiety. In *Anxiety* (ed. N. McNaughton and G. Andrews), pp. 124–33. Otago University Press, Dunedin.

Blanchard, R. J. and Blanchard, D. C. (1990b). Anti-predator defense as models of animal fear and anxiety. In *Fear and defence* (ed. P. F. Brain, S. Parmigiani, R. J. Blanchard, and D. Mainardi), pp. 89–108. Harwood Academic, Chur.

Blanchard, R. J. and Blanchard, D. C. (1994). Environmental targets and sensorimotor systems in aggression and defence. In *Ethology and psychopharmacology* (ed. S. J. Cooper and C. A. Hendrie), pp. 133–57. Wiley.

Blanchard, R. J. and Blanchard, D. C. (1996). Defensive quiescence to social stress: relationship to other 'immobile' defenses. *Society for Neuroscience Abstracts*, **22**, 1138.

Blanchard, R. J., Flannelly, K. J., and Blanchard, D. C. (1986). Defensive behaviors of laboratory and wild *Rattus norvegicus*. *Journal of Comparative Psychology*, **100**, 101–7.

Blanchard, R. J., Yudko, E. B., Rodgers, R. J., and Blanchard, D. C. (1993). Defense system psychopharmacology: an ethological approach to the pharmacology of fear and anxiety. *Behavioural Brain Research*, **58**, 155–65.

Blanchard, R. J., Parmigiani, S., Bjornson, C., Masuda, C., Weiss, S. M., and Blanchard, D. C. (1995). Antipredator behaviour of Swiss–Webster mice in a visible burrow system. *Aggressive Behaviour*, **21**, 123–36.

Blanchard, R. J., Griebel, G., Henrie, J. A., and Blanchard, D. C. (1997). Differentiation of anxiolytic and panicolytic drugs by effects on rat and mouse defense test batteries. *Neuroscience and Biobehavioral Reviews*, **21**, 783–9.

Bliss, T. V. P. and Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetised rabbit following stimulation of the perforant path. *Journal of Physiology (London)*, **232**, 331–56.

Blizard, D. A. (1981). The Maudsley reactive and nonreactive strains: a North American perspective. *Behavior Genetics*, **11**, 469–89.

Blizard, D. A. and Liang, B. (1979). Plasma catecholamines under basal and stressful conditions in rat strains selectively bred for differences in response to stress. In *Catecholamines*: *basic and clinical frontiers* (ed. E. Usdin, I. J. Kopin, and J. Bachas), pp. 1795–7. Pergamon, New York.

Borlongan, C. V. and Watanabe, S. (1994). Failure to discriminate conspecifics in amygdaloid-lesioned mice. *Pharmacology, Biochemistry and Behavior*, **48**, 677–80.

Braestrup, C. and Nielsen, M. (1980). Benzodiazepine receptors. *Arzneimittelforschung*, **30**, 852–7.

Braestrup, C. and Squires, R. F. (1977). Specific benzodiazepine receptors in rat brain characterized by high-affinity [3H]-diazepam binding. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 3805–9.

Brawman-Mintzer, O., Lydiard, R. B., Emmanuel, N., Payeur, R., Johnson, M., Roberts, J., *et al.* (1993). Psychiatric comorbidity in patients with generalized anxiety disorder. *American Journal of Psychiatry*, **150**, 1216–18.

Brawman-Mintzer, O., Lydiard, R. B., Crawford, M. M., Emmanuel, N., Payeur, R., Johnson, M., *et al.* (1994). Somatic symptoms in generalized anxiety disorder with and without comorbid psychiatric disorders. *American Journal of Psychiatry*, **151**, 930–2.

Breese, C. R., Hampson, R. E., and Deadwyler, S. A. (1989). Hippocampal place cells: stereotypy and plasticity. *The Journal of Neuroscience*, **9**, 1097–111.

Bremner, J. D., Krystal, J. H., Southwick, S. M., and Charney, D. S. (1995a). Functional neuroanatomical correlates of the effects of stress on memory. *Journal of Traumatic Stress*, **8**, 527–53.

Bremner, J. D., Randall, P., Scott, T. M., Bronen, R. A., Seibyl, J. P., Southwick, S. M., *et al.* (1995b). MRI-based measurement of hippocampal volume in patients with combat-related posttraumatic stress disorder. *American Journal of Psychiatry*, **152**, 973–81.

Bremner, J. D., Randall, P., Vermetten, E., Staib, L., Bronen, R. A., Mazure, C., *et al.* (1997). Magnetic resonance imaging-based measurement of hippocampal volume in post-traumatic stress disorder related to childhood physical and sexual abuse—a preliminary report. *Biological Psychiatry*, **41**, 23–32.

Brito, G. N. O. and Brito, L. S. O. (1990). Septohippocampal system and the prelimbic sector of frontal cortex: a neuropsychological battery analysis in the rat. *Behavioural Brain Research*, **36**, 127–46.

Broadbent, D. E. (1971). *Decision and stress*. Academic Press, London.

Broadhurst, P. L. (1960). Applications of biometrical genetics to the inheritance of behaviour. In *Experiments in personality*, Vol. 1, *Psychogenetics and psychopharmacology* (ed. H. J. Eysenck), pp. 1–102. Routledge Kegan Paul, London.

Broadhurst, P. L. (1975). The Maudsley reactive and nonreactive strains of rats: a survey. *Behavior Genetics*, **5**, 299–319.

Broekkamp, C. L. E., Berendsen, H. H. G., Jenck, F., and Van Delft, A. M. L. (1989). Animal models for anxiety and response to serotonergic drugs. *Psychopathology*, **22**, (Suppl. 1), 2–12.

Brooks, V. B. (1979). Motor programs revisited. In *Posture and movement* (ed. R. E. Talbott and D. R. Humphrey), pp. 13–49. Raven Press, New York.

Brown, J. S. (1961). *The motivation of behavior*. McGraw-Hill, New York.

Brown, J. W. (1985). Frontal lobe syndromes. In *Handbook of clinical neurology: clinical neuropsychology* (ed. J. A. M. Frederiks), pp. 23–41. Elsevier.

Brown, M. F. (1992). Does a cognitive map guide choices in the radial-arm maze? *Journal of Experimental Psychology: Animal Behaviour Processes*, **18**, 56–66.

Brown, M. W. (1982). Effect of context on the response of single units recorded from the hippocampal region of behaviourally trained monkeys. In *Neuronal plasticity and memory formation* (ed. C. A. Marsan and H. Matthies), pp. 557–73. Raven Press, New York.

Brown, T. A., Barlow, D. H., and Liebowitz, M. R. (1994). The empirical basis of generalized anxiety disorder. *American Journal of Psychiatry*, **151**, 1272–80.

Brown, T. A., Chorpita, B. F., and Barlow, D. H. (1998). Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology*, **107**, 179–92.

Browne, T. R. and Perry, J. K. (1973). Benzodiazepines in the treatment of epilepsy. *Epilepsia*, **14**, 277–310.

Brush, F. R., Baron, S., Froehlich, J. C., Ison, J. R., Pelligrino, L. J., Phillips, D. S., *et al.* (1985). Genetic differences in avoidance learning by *rattus norvegicus*: escape/avoidance responding, sensitivity to electric shock, discrimination learning, and open-field behavior. *Journal of Comparative Psychology*, **99**, 60–73.

Brutkowski, S. (1965). Functions of prefrontal cortex in animals. *Physiological Reviews*, **45**, 721–46.

Büchel, C., Morris, J., Dolan, R. J., and Friston, K. J. (1998). Brain systems mediating aversive conditioning: an event-related fMRI study. *Neuron*, **20**, 947–57.

Buhusi, C. V., Gray, J. A., and Schmajuk, N. A. (1998). Perplexing effects of hippocampal lesions on latent inhibition: a neural network solution. *Behavioral Neuroscience*, **112**, 316–51.

Bunsey, M. and Eichenbaum, H. (1996). Conservation of hippocampal memory function in rats and humans. *Nature*, **379**, 255–7.

Butler, G., Fennell, M., Robson, P., and Gelder, H. (1991). Comparison of behaviour therapy and cognitive-behaviour therapy in the treatment of generalised anxiety disorder. *Journal of Consulting and Clinical Psychology*, **59**, 167–75.

Butler, R. W., Braff, D. C., Rausch, J. L., Jenkins, M. A., Sprock, J., and Geyer, M. A. (1990). Physiological evidence of exaggerated startle response in a subgroup of Vietnam veterans with combat-related PTSD. *American Journal of Psychiatry*, **147**, 1308–12.

Butters, N. and Cermak, L. S. (1975). Some analyses of amnesic syndromes in brain-damaged patients. In *The hippocampus*, Vol. 2, *Neurophysiology and behaviour* (ed. R. L. Isaacson and K. H. Pribram), pp. 377–409. Plenum Press, New York.

Butters, N., Delis, D. C., and Lucas, J. A. (1995). Clinical assessment of memory disorders in amnesia and dementia. *Annual Review of Psychology*, **46**, 493–523.

Buzsàki, G. and Haas, H. L. (1988). Hippocampal sharp waves: a physiological correlate of LTP. In *Synaptic plasticity in the hippocampus* (ed. G. Buzsàki and H. L. Haas), pp. 90–102. Springer, Heidelberg.

Cahill, L. and McGaugh, J. L. (1996). Modulation of memory storage. *Current Opinion in Neurobiology*, **6**, 237–42.

Cahusac, P. M. B., Miyashita, Y., and Rolls, E. T. (1989). Responses of hippocampal formation neurons in the monkey related to delayed spatial response and object–place memory tasks. *Behavioural Brain Research*, **33**, 229–40.

Caldaroni, B., Saavedra, C., Tartaglia, K. L., Wehner, J. M., Dudek, B. C., and Flaherty, L. (1997). Quantitative trait loci analysis affecting contextual conditioning in mice. *Nature Genetics*, **17**, 335–7.

Calder, A. J., Young, A. W., Rowland, D., Perrett, D. I., Hodges, J. R., and Etcoff, N. L. (1996). Facial emotion recognition after bilateral amygdala damage: differentially severe impairment of fear. *Cognitive Neuropsychology*, **13**, 699–740.

Campeau, S. and Davis, M. (1995a). Involvement of the central nucleus and basolateral complex of the amygdala in fear conditioning measured with fear-potentiated startle in rats trained concurrently with auditory and visual conditioned stimuli. *The Journal of Neuroscience*, **15**, 2301–11.

Campeau, S. and Davis, M. (1995b). Involvement of subcortical and cortical afferents to the lateral nucleus of the amygdala in fear conditioning measured with fear-potentiated startle in rats trained concurrently with auditory and visual conditioned stimuli. *The Journal of Neuroscience*, **15**, 2312–27.

Carli, M. and Samanin, R. (1992). 8-Hydroxy-2-(di-*N*-propylamino)tetralin impairs spatial learning in a water maze: role of postsynaptic $5HT_{1A}$ receptors. *British Journal of Pharmacology*, **105**, 720–6.

Carlson, N. R. (1980). *Physiology of behavior*. Allyn and Bacon, Boston.

Carter, C., Maddock, R., Zoglio, M., Lutrin, C., Jella, S., and Amsterdam, E. (1994). Panic disorder and chest pain: a study of cardiac stress scintigraphy patients. *American Journal of Cardiology*, **74**, 296–8.

Cattell, R. B. (1965). *The scientific analysis of personality*. Pelican, Harmondsworth.

Cattell, R. B. and Scheier, I. H. (1961). *The meaning and measurement of neuroticism and anxiety*. Ronald, New York.

Cermak, L. S. and Butters, N. (1972). The role of interference and encoding in the short-term memory deficits of Korsakoff patients. *Neuropsychologia*, **10**, 89–95.

Cermak, L. S., Butters, N., and Gerrein, J. (1973). The extent of the verbal encoding ability of Korsakoff patients. *Neuropsychologia*, **11**, 85–94.

Cermak, L. S., Butters, N., and McReines, J. (1974). Some analyses of the verbal encoding deficit of alcoholic Korsakoff patients. *Brain and Language*, **2**, 141–50.

Chambless, D. L., Foa, E. B., Groves, G. A., and Goldstein, A. J. (1979). Flooding with Brevital in the treatment of agoraphobia: counter-effective? *Behaviour Research and Therapy*, **17**, 243–52.

Chao, L. L. and Knight, R. T. (1995). Human prefrontal lesions increase distractibility to irrelevant sensory inputs. *NeuroReport*, **6**, 1605–10.

Chen, Y.-C., Thaler, D., Nixon, P. D., Stern, C. E., and Passingham, R. E. (1995). The functions of the medial premotor cortex. II. The timing and selection of learned movements. *Experimental Brain Research*, **102**, 461–73.

Chi, C. C. (1965). The effect of amobarbital sodium on conditioned fear as measured by the potentiated startle response in rats. *Psychopharmacologia*, **7**, 115–22.

Chiba, A. A., Kesner, R. P., and Reynolds, A. M. (1994). Memory for spatial location as a function of temporal lag in rats: role of hippocampus and medial prefrontal cortex. *Behavioral and Neural Biology*, **61**, 123–31.

Clark, D. M. (1986). A cognitive approach to panic. *Behaviour Research and Therapy*, **24**, 461–70.

Clark, D. M. (1988). A cognitive model of panic attacks. In *Panic: psychological perspectives* (ed. S. Rachman and J. Maser), pp. 71–90. Erlbaum, Hillsdale, NJ.

Clark, D. M. (1991). *Cognitive therapy for panic disorder*. Paper presented at the National Institutes of Health Consensus Development Conference on the Treatment of Panic Disorder.

Clark, D. M. (1996). Panic disorder: from theory to therapy. In *Frontiers of cognitive therapy* (ed. P. M. Salkovskis), pp. 318–44. Guildford Publications.

Clark, D. M. (1997). Panic disorder and social phobia. In *Science and practice of cognitive behaviour therapy* (ed. D. M. Clark and C. G. Fairburn), pp. 121–53. Oxford University Press.

Clark, D. M. and Ehlers, A. (1993). An overview of the cognitive theory and treatment of panic disorder. *Applied and Preventative Psychology*, **2**, 131–9.

Clark, D. M. and Wells, A. (1995). A cognitive model of social phobia. In *Social phobia: diagnosis, assessment and treatment* (ed. R. G. Heimberg, M. R. Liebowitz, D. A. Hope, and F. R. Schneier), pp. 69–93. Guildford, New York.

Clark, D. M., Salkovskis, P. M., Hackmann, A., Middleton, H., Anastasiades, P., and Gelder, M. (1994). A comparison of cognitive therapy, applied relaxation and imipramine in the treatment of panic disorder. *British Journal of Psychiatry*, **164**, 759–69.

Cloninger, C. R., Svrakic, D. M., and Przybecky, T. R. (1993). A psychobiological model of temperament and character. *Archives of General Psychiatry*, **50**, 975–90.

Cohen, N. J. and Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal memory system*. MIT Press, Cambridge, Mass.

Cole, B. J. and Hillmann, M. (1994). Effects of benzodiazepine receptor ligands on the performance of an operant delayed matching to position task in rats: opposite effects of FG 7142 and lorazepam. *Psychopharmacology (Berlin)*, **115**, 350–7.

Cole, B. J., Jones, G. H., and Turner, J. D. (1994). $5HT_{1A}$ receptor agonists improve the performance of normal and scopolamine-impaired rats in an operant delayed matching to position task. *Psychopharmacology (Berlin)*, **116**, 135–42.

Cole, J. C. and Rodgers, R. J. (1993). An ethological analysis of the effects of chlordiazepoxide and bretazenil (Ro 16-6028) in the murine elevated plus-maze. *Behavioural Pharmacology*, **4**, 573–80.

Colombo, M., Eickhoff, A. E., and Gross, C. G. (1993). The effects of inferior temporal and dorsolateral frontal lesions on serial-order behavior and visual imagery in monkeys. *Cognitive Brain Research*, **1**, 211–17.

Commissaris, R. L., Verbanac, J. S., Markovska, V. L., Altman, H. J., and Hill, T. J. (1996). Anxiety-like and depression-like behavior in Maudsley reactive (MR) and nonreactive (MNRA) rats. *Progress in Neuropsychopharmacology and Biological Psychiatry*, **20**, 491–501.

Coop, C. F., McNaughton, N., Warnock, K., and Laverty, R. (1990). Effects of ethanol and Ro 15–4513 in an electrophysiological model of anxiolytic action. *Neuroscience*, **35**, 669–74.

Corkin, S. (1968). Acquisition of motor skill after bilateral medial temporal-lobe excision. *Neuropsychologia*, **6**, 255–65.

Costa, P. T. and McCrae, R. R. (1985). *The NEO personality inventory manual*. Psychological Assessment Resources, Odessa, FL.

Cowley, D. S., Roy-Byrne, P. P., Radant, A., Ritchie, J. C., Greenblatt, D. J., Nemeroff, C. B., *et al.* (1995). Benzodiazepine sensitivity in panic disorder: effects of chronic alprazolam treatment. *Neuropsychopharmacology*, **12**, 147–57.

Crawley, J. N. and Corwin, R. L. (1994). *Biological actions of cholecystokinin*. Peptides, **15**, 731–55.

Criswell, H. E., Knapp, D. J., Overstreet, D. H., and Breese, G. R. (1994). Effects of ethanol, chlordiazepoxide, and MK-801 on performance in the elevated-plus maze and on locomotor activity. *Alcohol Clinical Experimental Research*, **18**, 596–601.

Crovetti, R., Lucchetti, F., Mancia, M., Mariotti, M., Porrini, M., Spinnler, P., *et al.* (1995). Food intake after amygdaloid lesion in rats. *Nutrition Research*, **15**, 565–70.

Curran, H. V. (1991). Benzodiazepines, memory and mood: a review. *Psychopharmacology*, **105**, 1–8.

Cutler, N. R., Sramek, J. J., Keppel Hesselink, J. M., Krol, A., Roeschen, J., Rickels, K., *et al.* (1993). A double-blind, placebo-controlled study comparing the efficacy and safety of ipsapirone versus lorazepam in patients with generalized anxiety disorder: a prospective multicenter trial. *Journal of Clinical Psychopharmacology*, **13**, 429–37.

Dalgleish, T. and Power, M. (ed.) (1999). *Handbook of cognition and emotion*. Wiley.

Danion, J. M., Weingartner, H., File, S. E., Jaffard, R., Sunderland, T., Tulving, E., *et al.* (1993). Pharmacology of human memory and cognition: illustrations from the effects of benzodiazepines and cholinergic drugs. *Journal of Psychopharmacology*, 7, 371–7.

Dantendorfer, K., Prayer, D., Kramer, J., Amering, M., Baischer, W., Berger, P., *et al.* (1996). High frequency of EEG and MRI brain abnormalities in panic disorder. *Psychiatry Research: Neuroimaging Section*, **68**, 41–53.

Daum, I. and Ackermann, H. (1994). Dissociation of declarative and nondeclarative memory after bilateral thalamic lesions: a case report. *International Journal of Neuroscience*, **75**, 153–65.

Daum, I., Channon, S., Polkey, C. E., and Gray, J. A. (1991). Classical conditioning after temporal lobe lesions in man: impairment in conditional discrimination. *Behavioral Neuroscience*, **105**, 396–408.

Davey, G. C. L. (1992). Classical conditioning and the acquisition of human fears and phobias: a review and synthesis of the literature. *Advances in Behavioural Research Therapy*, **14**, 29–66.

Davey, G. C. L. (1995). Preparedness and phobias: specific evolved associations or a generalized expectancy bias? *Behavioural and Brain Sciences*, **18**, 289–97.

Davidson, T. L. (1992). Support for configural association theory: now you see it, now you don't. *Hippocampus*, **2**, 90–1.

Davidson, T. L. and Jarrard, L. E. (1993). A role for hippocampus in the utilization of hunger signals. *Behavioral and Neural Biology*, **59**, 167–71.

Davidson, T. L., McKernan, M. G., and Jarrard, L. E. (1993). Hippocampal lesions do not impair negative patterning: a challenge to configural association theory. *Behavioral Neuroscience*, **107**, 227–34.

Davies, M. F., Onaivi, E. S., Chen, S.-W., Maguire, P. A., Tsai, N. F., and Loew, G. H. (1994). Evidence for central benzodiazepine receptor heterogeneity from behavior tests. *Pharmacology, Biochemistry and Behavior*, **49**, 47–56.

Davis, M. (1992a). The role of the amygdala in conditioned fear. In *The amygdala: neurobiological aspects of emotion, memory, and mental function* (ed. J. P. Aggleton), pp. 255–305. Wiley-Liss.

Davis, M. (1992b). The role of the amygdala in fear-potentiated startle: implications for animal models of anxiety. *Trends in Pharmacological Science*, **13**, 35–41.

Davis, M. (1992c). The role of the amygdala in fear and anxiety. *Annual Review of Neuroscience*, **15**, 353–75.

Davis, M. (1998). Are different parts of the extended amygdala involved in fear versus anxiety? *Biological Psychiatry*, **44**, 1239–47.

Davis, M., Falls, W. A., Campeau, S., and Kim, M. (1993). Fear-potentiated startle: a neural and pharmacological analysis. *Behavioural Brain Research*, **58**, 175–98.

Davis, M., Rainnie, D., and Cassell, M. (1994). Neurotransmission in the rat amygdala related to fear and anxiety. *Trends in Neuroscience*, **17**, 208–14.

Davis, N. M., Brookes, S., Gray, J. A., and Rawlins, J. N. P. (1981). Chlordiazepoxide and resistance to punishment. *Quarterly Journal of Experimental Psychology*, **33B**, 227–39.

Dawson, G. R. and Tricklebank, M. D. (1995). Use of the elevated plus maze in the search for novel anxiolytic agents. *Trends in Pharmacological Science*, **16**, 33–6.

Dawson, G. R., Crawford, S. P., Collinson, N., Iversen, S. D., and Tricklebank, M. D. (1995). Evidence that the anxiolytic-like effects of chlordiazepoxide on the elevated plus maze are confounded by increases in locomotor activity. *Psychopharmacology (Berlin)*, **118**, 316–23.

De Souza Silva, M. A. and Tomaz, C. (1995). Amnesia after diazepam infusion into basolateral but not central amygdala of *Rattus norvegicus*. *Neuropsychobiology*, **32**, 31–6.

Deacon, R. M. J. and Rawlins, J. N. P. (1995). Serial position effects and duration of memory for nonspatial stimuli in rats. *Journal of Experimental Psychology: Animal Behaviour*, **21**, 285–92.

Deakin, J. F. W. (1993). A review of clinical efficacy of $5HT_{1A}$ agonists in anxiety and depression. *Journal of Psychopharmacology*, 7, 283–9.

Deakin, J. F. W. and Graeff, F. G. (1991). 5HT and mechanisms of defence. *Journal of Psychopharmacology*, **5**, 305–15.

DeFries, J. C. and Hegman, J. P. (1970). Genetic analysis of open-field behaviour. In *Contributions to behavior-genetic analysis: the mouse as a prototype* (ed. G. Lindzey and D. D. Thiessen), pp. 23–56. Appleton-Century-Crofts, New York.

DeFries, J. C., Wilson, J. R., and McClearn, G. E. (1970). Open-field behaviour in mice: selection response and situational generality. *Behavior Genetics*, **1**, 195–211.

Dellemijn, P. L. I. and Fields, H. L. (1994). Do benzodiazepines have a role in chronic pain management? *Pain*, **57**, 137–52.

Depue, R. and Collins, P. (1999). Neurobiology of the structure of personality: dopamine facilitation of incentive, motivation, and extraversion. *Behavioural and Brain Sciences*, **22**, 491–569.

Deuel, R. and Mishkin, M. (1977). Limbic and prefrontal contributions to somesthetic learning in monkeys. *Brain Research*, **132**, 521–35.

Devenport, L., Stidhan, J., and Hale, R. (1989). Ethanol and spatial localisation. *Behavioral Neuroscience*, **103**, 1259–66.

Diamond, A. (1995). Evidence of robust recognition memory early in life even when assessed by reaching behaviour. *Journal of Experimental Child Psychology*, **59**, 419–56.

Dickinson, A. (1974). Response suppression and facilitation by aversive stimuli following septal lesions in rats: a review and model. *Physiological Psychology*, **2**, 444–56.

Dickinson, A. (1975). Suppressive and enhancing effects of footshock on food-reinforced operant responding following septal lesions in rats. *Journal of Comparative and Physiological Psychology*, **88**, 851–61.

Dickinson, A. (1980). *Contemporary animal learning theory*. Cambridge University Press.

Dickinson, A. and Pearce, J. M. (1977). Inhibitory interactions between appetitive and aversive stimuli. *Psychological Bulletin*, **84**, 690–711.

DiScenna, P. G., Nowicky, A. V., and Teyler, T. J. (1994). The development of $GABA_B$-mediated activity in the rat dentate gyrus. *Developmental Brain Research*, 77, 295–8.

Doble, A. and Martin, I. L. (1992). Multiple benzodiazepine receptors: no reason for anxiety. *Trends in Pharmacological Science*, **13**, 76–81.

Dobson, K. S. and Block, L. (1988). Historical and philosophical bases of the cognitive-behavioural therapies. In *Handbook of cognitive-behavioural therapies* (ed. K. S. Dobson), pp. 3–38. The Guildford Press, New York.

Donovick, P. J., Burright, R. G., and Fink, E. A. (1979). Discrimination behavior of rats with septal lesions in a cue addition–deletion paradigm. *Physiology and Behavior*, **22**, 125–31.

Douglas, R. J. (1967). The hippocampus and behavior. *Psychological Bulletin*, **67**, 416–42.

Drewe, E. A. (1975a). An experimental investigation of Luria's theory on the effects of frontal lobe lesions in man. *Neuropsychologia*, **13**, 421–9.

Drewe, E. A. (1975b). Go–no go learning after frontal lobe lesions in humans. *Cortex*, **11**, 8–16.

Drewett, R. F., Gray, J. A., James, D. T. D., McNaughton, N., Valero, I., and Dudderidge, H. J. (1977). Sex and strain differences in septal driving of the hippocampal theta rhythm as a function of frequency: effects of gonadectomy and gonadal hormones. *Neuroscience*, **2**, 1033–41.

Eaves, L. and Meyer, J. (1994). Locating human quantitative trait loci: guidelines for the selection of sibling pairs for genotyping. *Behavior Genetics*, **2**, 443–55.
Ehlers, A. and Breuer, P. (1996). How good are patients with panic disorder at perceiving their heartbeats. *Biological Psychiatry*, **42**, 165–82.
Eibl-Eibesfeldt, I. and Sutterlin, C. (1990). Fear, defense and aggression in animals and man: some ethological perspectives. In *Fear and defense* (ed. P. F. Brain, S. Parmigian, R. Blanchard, R. Mainardi, and D. Harwood), pp. 381–408. Academic, Chur.
Eichenbaum, H. (1996). Is the rodent hippocampus just for 'place'? *Current Opinion in Neurobiology*, **6**, 187–95.
Eichenbaum, H., Shedlack, K. J., and Eckmann, K. W. (1980). Thalamocortical mechanisms in odor-guided behavior. I. Effects of lesions of the mediodorsal thalamic nucleus and frontal cortex on olfactory discrimination in the rat. *Brain Behaviour and Evolution*, **17**, 255–75.
Eichenbaum, H., Kuperstein, M., Fagan, A., and Nagode, J. (1987). Cue-sampling and goal-approach correlates of hippocampal unit activity in rats performing an odor-discrimination task. *The Journal of Neuroscience*, 7, 716–32.
Eichenbaum, H., Steward, C., and Morris, R. G. M. (1990). Hippocampal representation in place learning. *The Journal of Neuroscience*, **10**, 3531–42.
Eichenbaum, H., Otto, T., and Cohen, N. J. (1994). Two functional components of the hippocampal memory system. *Behavioural and Brain Sciences*, **17**, 449–518.
Eichenbaum, H., Otto, T., and Cohen, N. J. (1996). The hippocampal system: dissociating its functional components and recombining them in the service of declarative memory. *Behavioural and Brain Sciences*, **19**, 772–6.
Ekman, P. (1982). *Emotion in the human face*. Cambridge University Press.
Ekman, P. and Friesen, W. V. (1986). A new pan-cultural facial expression of emotion. *Motivation and Emotion*, **10**, 159–68.
Elliott, K. and Whelan, J. (eds.) (1978). *Functions of the septo-hippocampal system*, Ciba Foundation Symposium 58 (New Series). Elsevier, Amsterdam.
Ellis, A. (1962). *Reason and emotion in psychotherapy*. Citadel Press, Secaucus, NJ.
Evans, M. D., Hollon, S. D., DeRubeis, R. J., Piasecki, J. M., Grove, W. M., and Tuason, V. B. (1992). Differential relapse following therapy and pharmacotherapy for depression. *Archives of general psychiatry*, **48**, 80–8.
Eysenck, H. J. (ed.) (1981). *A model for personality*. Springer, New York.
Eysenck, H. J. (1957). *The dynamics of anxiety and hysteria*. Praeger, New York.
Eysenck, H. J. (1960). Classification and the problem of diagnosis. In *Handbook of abnormal psychology* (ed. H. J. Eysenck), pp. 1–31. Pitman, London.
Eysenck, H. J. (1967). *The biological basis of personality*. Thomas, Springfield, IL.
Eysenck, H. J. (1977). *You and neurosis*. Temple Smith, London.
Eysenck, H. J. (1979). The conditioning model of neurosis. *Behavioural and Brain Sciences*, **2**, 155–66.
Eysenck, H. J. and Eysenck, S. B. G. (1969). *The structure and measurement of personality*. Routledge and Kegan Paul, London.
Eysenck, H. J. and Eysenck, S. B. G. (1976). *Psychoticism as a dimension of personality*. Hodder and Stoughton, London.
Eysenck, H. J. and Rachman, S. (1965). *The causes and cures of neurosis*. Knapp, San Diego.
Eysenck, M. W. (1977). *Human memory: theory, research and individual differences*. Pergamon, Oxford.
Eysenck, M. W. (1992a). *Anxiety: the cognitive perspective*. Erlbaum, Hillsdale, NJ.

Eysenck, M. W. (1992b). The nature of anxiety. In *Handbook of individual differences; biological perspectives* (ed. A. Gale and M. W. Eysenck), pp. 157–78. Wiley.

Eysenck, M. W. and Byrne, A. (1992). Anxiety and susceptibility to distraction. *Personality and Individual Differences*, **13**, 793–8.

Eysenck, M. W. and Calvo, M. G. (1992). Anxiety and performance; the processing efficiency theory. *Cognition and Emotion*, **6**, 409–34.

Eysenck, M. W. and Eysenck, M. C. (1979). Memory scanning, introversion–extraversion and levels of processing. *Journal of Research in Personality*, **13**, 305–15.

Facklam, M., Schoch, P., and Haefely, W. E. (1992). Relationship between benzodiazepine receptor occupancy and potentiation of aminobutyric acid-stimulated chloride flux *in vitro* of four ligands of differing intrinsic efficacies. *Journal of Pharmacology and Experimental Therapeutics*, **261**, 1106–12.

Fairburn, C. G., Jones, R., Peveler, R. C., Carr, S. J., Solomon, R. A., O'Connor, M. E., *et al.* (1991). Three psychological treatments for bulimia nervosa: a comparative trial. *Archives of General Psychiatry*, **48**, 463–9.

Falls, W. A. and Davis, M. (1995). Lesions of the central nucleus of the amygdala block conditioned excitation, but not conditioned inhibition of fear as measured with the fear-potentiated startle effect. *Behavioral Neuroscience*, **109**, 379–87.

Fanselow, M. S. (1991). The midbrain periaqueductal gray as a coordinator of action in response to fear and anxiety. In *The midbrain periaqueductal gray matter* (ed. A. Depaulis and R. Bandler), pp. 151–73. Plenum Press, New York.

Fava, G. A., Grandi, S., Rafanelli, C., Saviotti, F. M., Ballin, M., and Pesarin, F. (1993). Hostility and irritable mood in panic disorder with agoraphobia. *Journal of Affective Disorders*, **29**, 213–17.

Feder, R. and Ranck, J. B., Jr. (1973). Studies on single neurons in dorsal hippocampal formation and septum in unrestrained rats. II. Hippocampal slow waves and theta cell firing during bar pressing and other behaviours. *Experimental Neurology*, **41**, 532–55.

Feighner, J. P. (1987). Buspirone in the long-term treatment of generalized anxiety disorder. *The Journal of Clinical Psychiatry*, **48**, 3–6.

Feldon, J. (1977). The effects of anti-anxiety drugs and selective lesions of the septo-hippocampal system on behavioural responses to non-reward and punishment. Unpublished D.Phil. thesis. University of Oxford.

Feldon, J. and Gray, J. A. (1981). The partial reinforcement extinction effect after treatment with chlordiazepoxide. *Psychopharmacology*, **73**, 269–75.

Feldon, J., Guillamon, A., Gray, J. A., De Wit, H., and McNaughton, N. (1979). Sodium amylobarbitone and responses to nonreward. *Quarterly Journal of Experimental Psychology*, **31**, 19–50.

Fendt, M., Koch, M., and Schnitzler, H.-U. (1994). Amygdaloid noradrenaline is involved in the sensitization of the acoustic startle response in rats. *Pharmacology, Biochemistry and Behavior*, **48**, 307–14.

Fernandez-Teruel, A., Escorihuela, R. M., Gray, J. A., Aguilar, R., Gil, L., Gimenez-Llort, L., Tobena, A., Bhomra, A., Nicod, A., Mott, R., Driscoll, P., Dawson, G. R., and Flint, J. (2002). A quantitative trait locus influencing anxiety in the laboratory rat, *Genome Research*, **12**, 618–26.

File, S. E. (1986). Intrinsic actions of the benzodiazepine receptor antagonist Ro 15–1788. *Psychopharmacology*, **88**, 1–11.

File, S. E. (1992). Behavioural detection of anxiolytic action. In *Experimental approaches to anxiety and depression* (ed. J. M. Elliott, D. J. Heal, and C. A. Marsden), pp. 25–44. Wiley, London.

File, S. E. and Cooper, S. J. (1985). Benzodiazepines and behavior. *Neuroscience and Biobehavioral Reviews*, **9**, 1–3.

Flint, J., Corley, R., DeFries, J. C., Fulker, D. W., Gray, J. A., Miller, S., *et al.* (1995). A simple genetic basis for a complex psychological trait in laboratory mice. *Science*, **269**, 1432–5.

Floderus-Myhred, B., Pedersen, N., and Rasmuson, I. (1980). Assessment of heritability for personality, based on a short-form of the Eysenck Personality Inventory: a study of 12,898 twin pairs. *Behavior Genetics*, **10**, 153–62.

Flynn, J. P. (1967). The neural basis of aggression in cats. In *Neurophysiology and emotion* (ed. D. C. Glass), pp. 40–60. Rockefeller University Press, New York.

Flynn, J. P. (1969). Neural aspects of attack behaviour in cats. *Annals of the New York Academy of Sciences*, **159**, 1008–12.

Foster, T. C., Christian, E. P., Hampson, R. E., Campbell, K. A., and Deadwyler, S. A. (1987). Sequential dependencies regulate sensory evoked responses of single units in the rat hippocampus. *Brain Research*, **408**, 86–96.

Foster, T. C., Castro, C. A., and McNaughton, B. L. (1989). Spatial selectivity of rat hippocampal neurons: dependence on preparedness for movement. *Science*, **244**, 1580–2.

Fowles, D. C. (1980). The three arousal model: implications of Gray's two-factor learning theory for heart rate, electrodermal activity and psychopathy. *Psychophysiology*, **17**, 87–104.

Frankland, P. W. and Yeomans, J. S. (1995). Fear-potentiated startle and electrically evoked startle mediated by synapses in rostrolateral midbrain. *Behavioral Neuroscience*, **109**, 669–80.

Franklin, J. A. (1990). Behavioural treatment for panic disorder. In *Anxiety* (ed. N. McNaughton and G. Andrews), pp. 84–91. University of Otago, Dunedin.

Fredrikson, M. and Öhman, A. (1979). Cardiovascular and electrodermal responses conditioned to fear-relevant stimuli. *Psychophysiology*, **16**, 1–7.

Fredrikson, M., Wik, G., Annas, P., Ericson, K., and Stone-Elander, S. (1995). Functional neuroanatomy of visually elicited simple phobic fear: additional data and theoretical analysis. *Psychophysiology*, **32**, 43–8.

Freed, D. M., Corkin, S., and Cohen, N. J. (1987). Forgetting in H.M.: a second look. *Neuropsychologia*, **25**, 461–71.

Freund, T. F. and Antal, M. (1988). GABA-containing neurons in the septum control inhibitory interneurons in the hippocampus. *Nature*, **336**, 170–3.

Friend, P. and Andrews, G. (1990). Agoraphobia without panic attacks. In *Anxiety* (ed. N. McNaughton and G. Andrews), pp. 67–73. University of Otago, Dunedin.

Frisby, J. P. (1979). *Seeing*. Oxford University Press.

Frith, C. D., Johnstone, E. C., Owens, D., and McPherson, K. (1979). The phenomenological, pharmacological and psychophysiological characteristics of neurotic outpatients to whom no diagnosis has been applied. In *Neuropsychopharmacology* (ed. B. Saletu, P. Berner, and L. Hollister), pp. 473–7. Pergamon, Oxford.

Frysztak, R. J. and Neafsey, E. J. (1994). The effect of medial frontal cortex lesions on cardiovascular conditioned emotional responses in the rat. *Brain Research*, **643**, 181–93.

Fulker, D. W., Cherny, S. S., and Cardon, L. R. (1995). Multipoint interval mapping of quantitative trait loci using sib pairs. *American Journal of Human Genetics*, **56**, 1224–33.

Funahashi, S. and Kubota, K. (1994). Working memory and prefrontal cortex. *Neuroscience Research*, **21**, 1–11.

Funahashi, S., Inoue, M., and Kubota, K. (1993). Delay-related activity in the primate prefrontal cortex during sequential reaching tasks with delay. *Neuroscience Research*, **18**, 171–5.

Furmark, T., Tillfors, M., Marteinsdottir, I., Fischer, H., Pissiota, A., Langstrom, B., and Fredrikson, M. (2002). Common changes in cerebral blood flow in patients with social phobia treated with citalopram or cognitive–behavioral therapy. *Archives of General Psychiatry*, **59**, 425–33.

Fuster, J. M. (1990). Behavioral electrophysiology of the prefrontal cortex of the primate. In *Progress in brain research* (ed. H. B. M. Uylings, C. G. Van Eden, J. P. C. De Bruin, M. A. Corner, and M. G. P. Feenstra), pp. 313–24. Elsevier.

Fuster, J. M. (1995). *Memory in the cerebral cortex*. MIT Press, Cambridge.

Fuxe, K., Agnati, L. F., Bolme, P., Hokfelt, T., Lidbrink, P., Ljungdahl, A., *et al.* (1975). The possible involvement of GABA mechanisms in the action of benzodiazepines on central catecholamine neurons. *Advances in Biochemistry and Pharmacology*, **14**, 45–61.

Gabrieli, J. D. E. (1995). A systematic view of human memory processes. *Journal of the International Neuropsychological Society*, **1**, 115–18.

Gaffan, D. (1972). Loss of recognition memory in rats with lesions of the fornix. *Neuropsychologia*, **10**, 327–41.

Gaffan, D. (1974). Recognition impaired and association intact in the memory of monkeys after transection of the fornix. *Journal of Comparative and Physiological Psychology*, **86**, 1100–9.

Gaffan, D. (1977a). Monkeys' recognition memory for complex pictures and the effect of fornix transection. *Quarterly Journal of Experimental Psychology*, **29**, 505–14.

Gaffan, D. (1977b). Recognition memory after short retention intervals in fornix-transected monkeys. *Quarterly Journal of Experimental Psychology*, **29**, 577–88.

Gaffan, D. (1994). Scene-specific memory for objects: a model of episodic memory impairment in monkeys with fornix transection. *Cognitive Neuroscience*, **6**, 305–20.

Gastaut, H. and Lammers, H. J. (1961). *Anatomie du Rhinencéphale*. Masson et Cie, Paris.

Gaykema, R. (1992). *The basal forebrain cholinergic system: organization of connections and long-term effects of lesions in the rat.* Unpublished PhD thesis, University of Groningen.

Gebhard, R., Zilles, K., Schleicher, A., Everitt, B. J., Robbins, T. W., and Divac, I. (1995). Parcellation of the frontal cortex of the New World monkey *Callithrix jacchus* by eight neurotransmitter-binding sites. *Anatomical Embryology (Berlin)*, **191**, 509–17.

Gehlert, D. R. (1994). Subtypes of receptors for neuropeptide Y: implications for the targeting of therapeutics. *Life Sciences*, **55**, 551–62.

Gelder, M. G., Bancroft, J. H. J., Gath, D. H., Johnston, D. W., Mathews, A. M., and Shaw, P. M. (1973). Specific and non-specific factors in behaviour therapy. *British Journal of Psychiatry*, **123**, 445–62.

George, F. and Johnson, L. (1985). *Purposive behaviour and teleological explanations*. Gordon and Breach, New York.

Gershenfeld, H. K., Neumann, P. E., Mathis, C., Crawley, J. N., Li, X., and Paul, S. M. (1997). Mapping quantitative trait loci for open-field behaviour in mice. *Behavior Genetics*, **27**, 201–10.

Givens, B. (1996a). Behavioral correlates of single units in the medial septal area: the effect of ethanol. *Neuroscience*, **71**, 417–27.

Givens, B. (1996b). Stimulus-evoked resetting of the dentate theta rhythm: relation to working memory. *NeuroReport*, **8**, 159–63.

Glaser, T. (1994). Ipsapirone, a potent and selective $5HT_{1A}$ receptor ligand with anxiolytic and antidepressant properties. *Correlates in Pharmacostructures*, **13**, 430–9.

Glassman, R. B., Garvey, K. J., Elkins, K. M., Kasal, K. L., and Couillard, N. L. (1994). Spatial working memory score of humans in a large radial maze, similar to published score of rats, implies capacity close to the magical number 7 ± 2. *Brain Research Bulletin*, **34**, 151–9.

Gloaguen, V., Cottraux, J., Cucherat, M., and Blackburn, I. M. (1998). A meta-analysis of the effects of cognitive therapy in depressed patients. *Journal of Affective Disorders*, **49**, 59–72.

Goeders, N. E., Ritz, M. C., and Kuhar, M. J. (1988). Buspirone enhances benzodiazepine receptor binding in vivo. *Neuropharmacology*, **27**, 275–80.

Goel, V. and Grafman, J. (1995). Are the frontal lobes implicated in 'planning' functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia*, **33**, 623–42.

Goisman, R. M., Warshaw, M. G., Peterson, L. G., Rogers, M. P., Cuneo, P., Hunt, M. F., *et al.* (1994). Panic, agoraphobia, and panic disorder with agoraphobia: data from a multicenter anxiety disorders study. *Journal of Nervous and Mental Diseases*, **182**, 72–9.

Goisman, R. M., Warshaw, M. G., Steketee, G. S., Fierman, E. J., Rogers, M. P., Goldenberg, I., *et al.* (1995). DSM-IV and the disappearance of agoraphobia without a history of panic disorder: new data on a controversial diagnosis. *American Journal of Psychiatry*, **152**, 1438–43.

Goldberg, D. P. and Huxley, P. (1992). *Common mental disorders: a biosocial model.* Routledge, London.

Goldberg, G. (1985). Supplementary motor area structure and function: review and hypotheses. *Behavioural and Brain Science*s, **8**, 567–616.

Goldberg, H. L. (1990). Modern treatment modalities in anxiety. In *Anxiety* (ed. N. McNaughton and G. Andrews), pp. 102–7. University of Otago, Dunedin.

Goldberg, H. L. and Finnerty, R. J. (1979). Comparative efficacy of buspirone and diazepam in the treatment of anxiety. *American Journal of Psychiatry*, **136**, 1184–7.

Goldman-Rakic, P. S. (1992). Working memory and the mind. *Scientific American*, **267**, 73–9.

Gorenstein, E. E. and Newman, J. P. (1969). Disinhibitory psychopathology: a new perspective and a model for research. *Psychological Review*, **87**, 301–15.

Gorman, J. M., Liebowitz, M. R., Fyer, A. J., and Stein, J. (1989). A neuroanatomical hypothesis for panic disorder. *American Journal of Psychiatry*, **146**, 148–61.

Gothard, K. M., Skaggs, W. E., Moore, K. M., and McNaughton, B. L. (1996). Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task. *The Journal of Neuroscience*, **16**, 823–35.

Graeff, F. G. (1991). Neurotransmitters in the dorsal periaqueductal grey and animal models of panic anxiety. In *New concepts in anxiety* (ed. M. Briley and S. E. File), pp. 288–312. MacMillan Press.

Graeff, F. G. (1994). Neuroanatomy and neurotransmitter regulation of defensive behaviors and related emotions in mammals. *Brazilian Journal of Medical and Biological Research*, **27**, 811–29.

Graeff, F. G., Viana, M. B., and Tomaz, C. (1993). The elevated T maze, a new experimental model of anxiety and memory: effect of diazepam. *Brazilian Journal of Medical and Biological Research*, **26**, 67–70.

Graeff, F. G., Guimaraes, F. S., De Andrade, T. G. C. S., and Deakin, J. F. W. (1996). Role of 5HT in stress, anxiety and depression. *Pharmacology, Biochemistry and Behavior,* **54**, 129–41.

Grafman, J. (1995). Similarities and distinctions among current models of prefrontal cortical functions. *Annals of the New York Academy of Sciences*, **769**, 337–68.

Granon, S., Vidal, C., Thinus-Blanc, C., Changeux, J.-P., and Poucet, B. (1994). Working memory, response selection, and effortful processing in rats with medial prefrontal lesions. *Behavioral Neuroscience*, **108**, 883–91.

Gray, J. A. (1964). Strength of the nervous system and levels of arousal: a reinterpretation. In *Pavlov's typology* (ed. J. A. Gray), pp. 289–366. Pergamon, Oxford.

Gray, J. A. (1967). Disappointment and drugs in the rat. Advancement of *Science*, **23**, 595–605.

Gray, J. A. (1969). Sodium amobarbital and effects of frustrative nonreward. *Journal of Comparative and Physiological Psychology*, **69**, 55–64.

Gray, J. A. (1970a). Sodium amobarbital, the hippocampal theta rhythm and the partial reinforcement extinction effect. *Psychological Review*, 77, 465–80.

Gray, J. A. (1970b). The psychophysiological basis of introversion–extraversion. *Behaviour Research and Therapy*, **8**, 249–66.

Gray, J. A. (1971). *The psychology of fear and stress*. Weidenfeld and Nicolson, London.

Gray, J. A. (1972). The structure of the emotions and the limbic system. In *Physiology, emotion and psychosomatic illness*, Ciba Foundation Symposium 8 (New Series) (ed. R. Porter and J. Knight), pp. 87–130. Associated Scientific, Amsterdam.

Gray, J. A. (1975). *Elements of a two-process theory of learning*. Academic Press, London.

Gray, J. A. (1976). The behavioural inhibition system: a possible substrate for anxiety. In *Theoretical and experimental bases of behaviour modification* (ed. M. P. Feldman and A. M. Broadhurst), pp. 3–41. Wiley, London.

Gray, J. A. (1977). Drug effects on fear and frustration: possible limbic site of action of minor tranquilizers. In *Handbook of psychopharmacology, volume 8. Drugs, neurotransmitters and behaviour* (ed. L. L. Iversen, S. D. Iversen, and S. H. Snyder), pp. 433–529. Plenum Press, New York.

Gray, J. A. (1979a). Is there any need for conditioning in Eysenck's conditioning model of neurosis? *Behavioural and Brain Science*s, **2**, 169–71.

Gray, J. A. (1979b). *Pavlov*. Fontana, London.

Gray, J. A. (1981). A critique of Eysenck's theory of personality. In *A model for personality* (ed. H. J. Eysenck), pp. 246–76. Springer, New York.

Gray, J. A. (1982). *The neuropsychology of anxiety: an enquiry into the functions of the septo-hippocampal system*. Oxford University Press.

Gray, J. A. (1987a). Interactions between drugs and behaviour therapy. In *Theoretical foundations of behaviour therapy* (ed. H. J. Eysenck and I. Martin), pp. 433–50. Plenum Press, New York.

Gray, J. A. (1987b). *The psychology of fear and stress*. Cambridge University Press, London.

Gray, J. A. (1995). The contents of consciousness: a neuropsychological conjecture. *Behavioural and Brain Sciences*, **18**, 659–722.

Gray, J. A. (1999) But the schizophrenia connection . . . [Commentary] *Behavioural and Brain Sciences*, **22**, 523–4.

Gray, J. A. and McNaughton, N. (1983). Comparison between the behavioural effect of septal and hippocampal lesions: a review. *Neuroscience and Biobehavioral Reviews*, **7**, 119–88.

Gray, J. A. and Smith, P. T. (1969). An arousal-decision model for partial reinforcement and discrimination learning. In *Animal discrimination learning* (ed. R. Gilbert and N. S. Sutherland), pp. 243–72. Academic Press, London.

Gray, J. A., McNaughton, N., James, D. T. D., and Kelly, P. H. (1975). Effect of minor tranquillisers on hippocampal theta rhythm mimicked by depletion of forebrain noradrenaline. *Nature (London)*, **258**, 424–5.

Gray, J. A., Feldon, J., Rawlins, J. N. P., Owen, S., and McNaughton, N. (1978). The role of the septo-hippocampal system and its noradrenergic afferents in behavioural responses to nonreward. In *Functions of the septo-hippocampal system*, Ciba Foundation Symposium No. 58 (New Series) (ed. K. Elliott and J. Whelan), pp. 275–300. Elsevier, Amsterdam.

Gray, J. A., Owen, S., Davis, N. M., and Tsaltas, E. (1983). Psychological and physiological relations between anxiety and impulsivity. In *The biological basis of impulsivity and sensation seeking* (ed. M. Zuckerman), pp. 181–217. Erlbaum, Hillsdale, NJ.

Gray, J. A., Feldon, J., Rawlins, J. N. P., Hemsley, D. R., and Smith, A. D. (1991). The neuropsychology of schizophrenia. *Behavioral and Brain Sciences*, **14**, 1–20.

Gray, J. A., Buhusi, C. V., and Schmajuk, N. A. (1997). The transition from automatic to controlled processing. *Neural Networks*, **10**, 1257–68.

Gray, J. A., Grigoryan, G., Virley, D., Sinden, J. D., and Hodges, H. (2000). Conditionally immortalised multipotential and multifunctional neural stem cell lines as an approach to clinical transplantation. *Cell Transplantation*. (In press.)

Griebel, G. (1995). 5-Hydroxytryptamine-interacting drugs in animal models of anxiety disorders: more than 30 years of research. *Pharmacology and Therapeutics*, **65**, 319–95.

Grillon, C., Ameli, R., Foot, M., and Davis, M. (1993). Fear-potentiated startle: relationship to the level of state/trait anxiety in healthy subjects. *Biological Psychiatry*, **33**, 566–74.

Gross, C. G. (1994). How inferior temporal cortex became a visual area. *Cerebral Cortex*, **5**, 455–69.

Groves, P. M. and Thompson, R. F. (1970). Habituation: a dual-process theory. *Psychological Review*, 77, 419–59.

Guidotti, A., Toffano, G., and Costa, E. (1978). An endogenous protein modulates the affinity of GABA and benzodiazepine receptors in rat brain. *Nature (London)*, **275**, 553–5.

Guidotti, A., Baraldi, M., Schwartz, J. P., and Costa, E. (1979). Molecular mechanisms regulating the interactions between the benzodiazepines and GABA receptors in the central nervous system. *Pharmacology, Biochemistry and Behavior*, **10**, 803–7.

Gurvits, T. V., Shenton, M. E., Hokama, H., Ohta, H., Lasko, N. B., Gilbertson, M. W., *et al.* (1996). Magnetic resonance imaging study of hippocampal volume in chronic, combat-related posttraumatic stress disorder. *Biological Psychiatry*, **40**, 1091–9.

Habets, A. M. M. C., Lopes da Silva, F. H., and Mollevanger, W. J. (1980). An olfactory input to the hippocampus of the cat: field potential analysis. *Brain Research*, **182**, 47–64.

Hackmann, A., Surawy, C., and Clark, D. M. (1998). Seeing yourself through others' eyes: a study of spontaneously occurring images in social phobia. *Behavioural and Cognitive Psychotherapy*, **26**, 3–12.

Haefely, W. (1990a). Benzodiazepine receptor and ligands: structural and functional differences. In *Benzodiazepines: current concepts* (ed. I. Hindmarch, G. Beaumont, S. Brandon, and B. E. Leonard), pp. 1–18. Wiley.

Haefely, W. (1990b). The $GABA_A$-benzodiazepine receptor: biology and pharmacology. In *Handbook of anxiety: the neurobiology of anxiety* (ed. G. D. Burrows, M. Roth, and R. Noyes), pp. 165–87. Elsevier.

Haefely, W. (1991). Comparative pharmacology of benzodiazepine receptor ligands with differing intrinsic efficacy. In *Transmitter amino acid receptors: structures, transduction and models for drug development* (ed. E. A. Barnard and E. Costa), pp. 91–111. Thieme, New York.

Haefely, W. (1992). The role of GABA in anxiolytic/antidepressant drug action. In *Experimental approaches to anxiety and depression* (ed. J. M. Elliott, D. J. Heal, and C. A. Marsden), pp. 151–68. Wiley.

Haefely, W., Martin, J. R., and Schoch, P. (1990). Novel anxiolytics that act as partial agonists at benzodiazepine receptors. *Trends in Pharmacological Science*, **11**, 452–6.

Haefely, W., Facklam, M., Schoch, P., Martin, J. R., Bonetti, E. P., Moreau, J., *et al.* (1992). Partial agonists of benzodiazepine receptors for the treatment of epilepsy, sleep, and anxiety disorders. In *GABAergic synaptic transmission* (ed. G. Biggio, A. Concas, and E. Costa), pp. 379–94. Raven Press, New York.

Haefely, W. E. (1990). The $GABA_A$–benzodiazepine receptor complex and anxiety. In *Anxiety: psychobiological and clinical perspectives* (ed. N. Sartorius, V. Andreoli, G. Cassano, L. Eisenberg, P. Kielholz, P. Pancheri, *et al.*), pp. 23–36. Hemisphere, New York.

Hafner, J. and Marks, I. (1976). Exposure in vivo of agoraphobics: contributions of diazepam, group exposure, and anxiety evocation. *Psychological Medicine*, **6**, 71–88.

Hall, C. S. (1951). The genetics of behavior. In *Handbook of experimental psychology* (ed. S. S. Stevens), pp. 304–29. Wiley, New York.

Hall, G., Purves, D., and Bonardi, C. (1996). Contextual control of conditioned responding in rats with dorsal hippocampal lesions. *Behavioral Neuroscience*, **110**, 933–45.

Hampson, R. E., Byrd, D. R., Konstantopoulos, J. K., Bunn, T., and Deadwyler, S. A. (1996). Hippocampal place fields: relationship between degree of field overlap and cross-correlations within ensembles of hippocampal neurons. *Hippocampus*, **6**, 281–93.

Handley, S. L. (1995). 5-Hydroxytryptamine pathways in anxiety and its treatment. *Pharmacology and Therapeutics*, **66**, 103–48.

Handley, S. L., McBlane, J. W., Critchley, M. A. E., and Njung'e, K. (1993). Multiple serotonin mechanisms in animal models of anxiety: environmental, emotional and cognitive factors. *Behavioural Brain Research*, **58**, 203–10.

Hannesson, D. K. and Skelton, R. W. (1998). Recovery of spatial performance in the Morris water maze following bilateral transection of the fimbria/fornix in rats. *Behavioural Brain Research*, **90**, 35–56.

Harper, D. N., McLean, A. P., and Dalrymple-Alford, J. C. (1994). Forgetting in rats following medial septum or mammillary body damage. *Behavioral Neuroscience*, **108**, 691–702.

Harris, J. A. and Westbrook, R. F. (1995). Effects of benzodiazepine microinjection into the amygdala or periaqueductal gray on the expression of conditioned fear and hypoalgesia in rats. *Behavioral Neuroscience*, **109**, 295–304.

Harro, J., Vasar, E., and Bradwejn, J. (1993). CCK in animal and human research on anxiety. *Trends in Pharmacological Science*, **14**, 244–9.

Hawton, K., Salkovskis, P. M., Kirk, J., and Clark, D. M. (1989). *Cognitive behaviour therapy for psychiatric problems*. Oxford University Press.

Hayne, H. (1998). Age-related changes in deferred imitation: the development of declarative memory. Paper presented at the Annual Meeting of the International Society for Developmental Biology, Orleans, France.

Hayne, H., MacDonald, S., and Barr, R. (1997). Developmental changes in the specificity of memory over the second year of life. *Infant Behavior and Development*, **20**, 233–45.

Hebb, D. O. (1946). On the nature of fear. *Psychological Review*, **53**, 259–76.

Hebb, D. O. (1949). *The organization of behavior*. Wiley-Interscience, New York.
Heimberg, R. G., Dodge, C. S., Hope, D. A., Kennedy, C. R., and Zollo, L. J. (1990). Cognitive behavioural group treatment for social phobia: comparison with a credible placebo control. *Cognitive Therapy and Research*, **17**, 235–48.
Hendrie, C. A., Neill, J. C., Shepherd, J. K., and Dourish, C. T. (1993). The effects of CCK_A and CCK_B antagonists on activity in the black/white exploration model of anxiety in mice. *Physiology and Behavior*, **54**, 689–93.
Henke, P. G. (1977). Dissociation of the frustration effect and the partial reinforcement extinction effect after limbic lesions in rats. *Journal of Comparative and Physiological Psychology*, **91**, 1032–8.
Hensman, R., Guimaraes, F. S., Wang, M., and Deakin, J. F. W. (1991). Effects of ritanserin on aversive classical conditioning in humans. *Psychopharmacology*, **104**, 220–4.
Herremans, A. H. J., Hijzen, T. H., Olivier, B., and Slangen, J. L. (1995). Relation between stimulus discriminability and response bias in a delayed conditional discrimination task in the rat. *Journal of Neuroscience Methods*, **63**, 29–35.
Herz, R. S. and Cupchik, G. C. (1995). The emotional distinctiveness of odor-evoked memories. *Chemical Senses*, **20**, 517–28.
Hill, A. J. (1978). First occurrence of hippocampal spatial firing in a new environment. *Experimental Neurology*, **62**, 282–97.
Hill, A. J. and Best, P. J. (1981). Effects of deafness and blindness on the spatial correlates of hippocampal unit activity in the rat. *Experimental Neurology*, **74**, 204–17.
Hinde, R. A. (1982). *Ethology*. Fontana.
Hinde, R. A. (1998). *Animal behaviour*. McGraw-Hill, New York.
Hinton, G. E., Plaut, D. C., and Shallice, T. (1993). Simulating brain damage. *Scientific American*, **269**, 58–65.
Hirata, H. and Aston-Jones, G. (1994). A novel long-latency response of locus coeruleus neurons to noxious stimuli: mediation by peripheral C-fibers. *Journal of Neurophysiology*, **71**, 1752–61.
Hirsh, R. (1974). The hippocampus and contextual retrieval of information from memory: a theory. *Behavioural Biology*, **12**, 421–4.
Hitchcock, J. M. and Davis, M. (1991). Efferent pathway of the amygdala involved in conditioned fear as measured with the fear-potentiated startle paradigm. *Behavioral Neuroscience*, **105**, 826–42.
Hodges, H. and Green, S. (1986). Effects of chlordiazepoxide on cued radial maze performance in rats. *Psychopharmacology (Berlin)*, **88**, 460–6.
Hodges, H., Green, S., and Glenn, B. (1987). Evidence that the amygdala is involved in benzodiazepine and serotonergic effects on punished responding but not on discrimination. *Psychopharmacology (Berlin)*, **92**, 491–504.
Hodgson, R. and Rachman, S. (1972). The effects of contamination and washing in obsessional patients. *Behaviour Research and Therapy*, **10**, 111–17.
Hof, P. R., Mufson, E. J., and Morrison, J. H. (1995). Human orbitofrontal cortex: cytoarchitecture and quantitative immunohistochemical parcellation. *The Journal of Comparative Neurology*, **359**, 48–68.
Hofer, M. A. (1972). Physiological and behavioural processes in early maternal deprivation. In *Physiology, emotion and psychosomatic illness* (ed. R. Porter and J. Knight), Ciba Symposium No. 8 (New Series), pp. 175–86. Elsevier.
Holdstock, T. L. (1969). Autonomic reactivity following septal and amygdaloid lesions in white rats. *Physiology and Behavior*, **4**, 603–7.

Holdstock, T. L. (1970). Plasticity of autonomic functions in rats with septal lesions. *Neuropsychologia*, **8**, 147–60.

Holley, L. A., Turchi, J., Apple, C., and Sarter, M. (1995). Dissociation between the attentional effects of infusions of a benzodiazepine receptor agonist and an inverse agonist into the basal forebrain. *Psychopharmacology (Berlin)*, **120**, 99–108.

Hollon, S. D., Shelton, R. C., and Loosen, P. T. (1991). Cognitive therapy and pharmacotherapy for depression. *Journal of Consulting and Clinical Psychology*, **59**, 88–99.

Hollon, S. D., DeRubeis, R. J., Evans, M. D., Weimber, M. J., Garvey, M. J., Grove, W. M., *et al.* (1992). Cognitive therapy and pharmacotherapy for depression: singly and in combination. *Archives of General Psychiatry*, **49**, 774–81.

Holt, P. (1990). Panic disorder: some historical trends. In *Anxiety* (ed. N. McNaughton and G. Andrews), pp. 54–65. University of Otago, Dunedin.

Homskaya, E. D. (1964). Verbal regulation of the vegetative components of the orienting reflex in focal brain lesions. *Cortex*, **1**, 63–76.

Horn, G. and Hinde, R. A. (ed.) (1971). *Short-term changes in neural activity and behaviour*. Cambridge University Press.

Hosobuchi, Y. (1981). Periaqueductal gray stimulation in humans produces analgesia accompanied by elevation of β-endorphin and ACTH in ventricular CSF. *Modern Problems in Pharmacopsychiatry*, **17**, 109–22.

Hugdahl, K. and Öhman, A. (1977). Effects of instruction on acquisition and extinction of electrodermal responses to fear-relevant stimuli. *Journal of Experimental Psychology: Human Learning and Memory*, **3**, 608–18.

Hugdahl, K. and Öhman, A. (1980). Skin conductance conditioning to potentially phobic stimuli as a function of interstimulus interval and delay versus trace paradigm. *Psychophysiology*, **17**, 348–55.

Hugdahl, K., Fredrikson, M., and Öhman, A. (1977). 'Preparedness' and 'arousability' as determinants of electrodermal conditioning. *Behaviour Research and Therapy*, **15**, 345–53.

Hunt, P. S., Richardson, R., and Campbell, B. A. (1994). Delayed development of fear-potentiated startle in rats. *Behavioral Neuroscience*, **108**, 69–80.

Hussain, M. Z. (1971). Desensitization and flooding (implosion) in treatment of phobias. *American Journal of Psychiatry*, **127**, 1509–14.

Hygge, S. and Öhman, A. (1978). Modeling processes in the acquisition of fears in vicarious electrodermal conditioning to fear-relevant stimuli. *Journal of Personality and Social Psychology*, **36**, 271–9.

Insausti, R., Amaral, D. G., and Cowan, W. M. (1987). The entorhinal cortex of the monkey: II. Cortical afferents. *The Journal of Comparative Neurology*, **264**, 356–95.

Ishimaru, H., Takahashi, A., Ikarashi, Y., and Maruyama, Y. (1995). Immunohistochemical and neurochemical studies of hippocampal cholinergic neurones after ischaemia. *NeuroReport*, **6**, 557–60.

Ison, J. R., Glass, D. H., and Bohmer, H. M. (1966). Effects of sodium amytal on the approach to stimulus change. *Proceedings of the American Psychological Association*, **2**, 5–6.

Jacobson, S., Butters, N., and Tovsky, N. J. (1978). Afferent and efferent subcortical projections of behaviorally defined sectors of prefrontal granular cortex. *Brain Research*, **159**, 279–96.

Jakes, I. (1996). *Theoretical approaches to obsessive–compulsive disorder*. Cambridge University Press.

James, D. T. D., McNaughton, N., Rawlins, J. N. P., Feldon, J., and Gray, J. A. (1977). Septal driving of hippocampal theta rhythm as a function of frequency in the free-moving male rat. *Neuroscience*, **2**, 1007–17.

James, W. (1884). What is an emotion? *Mind*, **9**, 188–205.

Janis, L. S., Bishop, T. W., and Dunbar, G. L. (1994). Medial septal lesions in rats produce permanent deficits for strategy selection in a spatial memory task. *Behavioral Neuroscience*, **108**, 892–8.

Jann, M. (1988). Buspirone: an update on a unique anxiolytic agent. *Pharmacotherapy*, **8**, 100–16.

Jansen, D. M. and Frijda, N. H. (1994). Modulation of the acoustic startle response by film-induced fear and sexual arousal. *Psychophysiology*, **31**, 565–71.

Jarrard, L. E. (1975). Role of interference in retention by rats with hippocampal lesions. *Journal of Comparative and Physiological Psychology*, **89**, 400–8.

Jarrard, L. E. (1978). Selective hippocampal lesions: differential effects on performance by rats of spatial task with preoperative versus postoperative training. *Journal of Comparative and Physiological Psychology*, **92**, 1119–27.

Jarrard, L. E. (1983). Selective hippocampal lesions and behavior: effects of kainic acid lesions on performance of place and cue tasks. *Behavioral Neuroscience*, **97**, 873–89.

Jarrard, L. E. (1991). On the neural basis of the spatial mapping system: hippocampus versus hippocampal formation. *Hippocampus*, **2**, 236–9.

Jarrard, L. E, Feldon, J., Rawlins, J. N. P., Sinden, J. D., and Gray, J. A. (1986). The effects of intrahippocampal ibotenate on resistance to extinction after continuous or partial reinforcement. *Experimental Brain Research*, **61**, 519–30.

Jaycox, L. H. and Foa, E. B. (1998). Post-traumatic stress disorder. In *Comprehensive clinical psychology* (ed. A. S. Bellack and M. Hersen), pp. 499–517. Elsevier, Oxford.

Johnson, T. N., Rosvold, H. E., and Mishkin, M. (1968). Projections from behaviorally-defined sectors of the prefrontal cortex to the basal ganglia, septum and diencephalon of the monkey. *Experimental Neurology*, **21**, 20–34.

Johnston, A. L. and File, S. E. (1988). Effects of ligands for specific 5HT receptor subtypes in two animal tests of anxiety. In *Buspirone: a new introduction to the treatment of anxiety*, Royal Society of Medicine Services International Congress and Symposium Series No. 133 (ed. M. Lader), pp. 31–7. Royal Society of Medicine Services.

Johnston, D. and Gath, D. (1973). Arousal levels and attribution effects in diazepam-assisted flooding. *British Journal of Psychiatry*, **123**, 463–6.

Johnstone, E. C., Owens, D. G. C., Frith, C. D., McPherson, K., Dowie, C., Riley, G., *et al.* (1980). Neurotic illness and its response to anxiolytic and antidepressant treatment. *Psychological Medicine*, **10**, 321–8.

Jung, M. W. and McNaughton, B. L. (1993). Spatial selectivity of unit activity in the hippocampal granular layer. *Hippocampus*, **3**, 165–82.

Jung, M. W., Wiener, S. I., and McNaughton, B. L. (1994). Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *The Journal of Neuroscience*, **14**, 7347–56.

Kamin, L. J. (1968). 'Attention-like' processes in classical conditioning. In *Miami symposium on the prediction of behaviour: aversive stimulation* (ed. M. R. Jones), pp. 9–31. University of Miami Press.

Kamin, L. J., Brimer, C. J., and Black, A. H. (1963). Conditioned suppression as a monitor of the fear of the CS in the course of avoidance training. *Journal of Comparative and Physiological Psychology*, **56**, 497–501.

Kandel, E. R. and Hawkins, R. D. (1992). The biological basis of learning and individuality. *Scientific American*, **267**, 79–86.

Kant, G. J., Wylie, R. M., Vasilakis, A. A., and Ghosh, S. (1996). Effects of triazolam and diazepam on learning and memory as assessed using a water maze. *Pharmacology, Biochemistry and Behavior*, **53**, 317–22.

Kapczinski, F., Curran, H. V., Gray, J., and Lader, M. (1994). Flumazenil has an anxiolytic effect in simulated stress. *Psychopharmacology (Berlin)*, **114**, 187–9.

Katz, R. J., Landau, P. S., Lott, M., Bystritsky, A., Diamond, B., Hoehn-Saric, R., *et al.* (1993). Serotonergic ($5HT_2$) mediation of anxiety—therapeutic effects of serazepine in generalized anxiety disorder. *Biological Psychiatry*, **34**, 41–4.

Kaviani, H., Gray, J. A., Checkley, S. A., Kumari, V., and Wilson, G. D. (1999). Modulation of the human acoustic startle reflex by emotionally-toned film-clips. *International Journal of Psychophysiology*, **32**, 47–54.

Keck, P. E., Jr., McElroy, S. L., Tugrul, K. C., Bennett, J. A., and Smith, J. M. R. (1993). Antiepileptic drugs for the treatment of panic disorder. *Neuropsychobiology*, **27**, 150–3.

Kelley, A. E. and Domesick, V. B. (1982). The distribution of the projection from the hippocampal formation to the nucleus accumbens in the rat: an anterograde and retrograde-horseradish peroxidase study. *Neuroscience*, **7**, 2321–35.

Kellner, R., Rada, R. T., Andersen, T., and Pathak, D. (1979). The effects of chlordiazepoxide on self-rated depression, anxiety, and well-being. *Psychopharmacology*, **64**, 185–91.

Kelly, D., Walter, C. J. S., Mitchell-Heggs, N., and Sargant, W. (1972). Modified leucotomy assessed clinically, physiologically and psychologically at six weeks and eighteen months. *British Journal of Psychiatry*, **120**, 19–29.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., and Eaves, L. J. (1992a). Generalized anxiety disorder in women: a population-based twin study. *Archives of General Psychiatry*, **49**, 267–72.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., and Eaves, L. J. (1992b). The genetic epidemiology of phobias in women: the interrelationship of agoraphobia, social phobia, situational phobia and simple phobia. *Archives of General Psychiatry*, **49**, 273–81.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., and Eaves, L. J. (1992c). Major depression and generalized anxiety disorder. Same genes, (partly) different environments? *Archives of General Psychiatry*, **49**, 716–22.

Kesner, R. P. (1993). Paired associate learning in the rat: role of hippocampus, medial prefrontal cortex, and parietal cortex. *Psychobiology*, **21**, 183–92.

Kesner, R. P. (1996). Parallels between animals and humans in terms of the neural foundations of memory: implication for Alzheimer's disease. In *Alzheimer's and Parkinson's diseases* (ed. H. J. Altman and B. N. Altman), pp. 209–20. Plenum Press, New York.

Kesner, R. P. and Beers, D. R. (1988). Dissociation of data-based and expectancy-based memory following hippocampal lesions in rats. *Behavioral and Neural Biology*, **50**, 46–60.

Kesner, R. P. and DeSpain, M. J. (1988). Correspondence between rats and humans in the utilization of retrospective and prospective codes. *Animal Learning and Behaviour*, **16**, 299–302.

Kesner, R. P. and Holbrook, T. (1987). Dissociation of item and order spatial memory in rats following medial prefrontal cortex lesions. *Neuropsychologia*, **25**, 653–64.

Kesner, R. P. and Novak, J. M. (1982). Serial position curve in rats: role of the dorsal hippocampus. *Science*, **218**, 173–5.

Kesner, R. P., Chiba, A. A., and Jackson-Smith, P. (1994a). Rats do show primacy and recency effects in memory for lists of spatial locations: a reply to Gaffan. *Animal Learning and Behaviour*, **22**, 214–18.

Kesner, R. P., Hopkins, R. O., and Fineman, B. (1994b). Item and order dissociation in humans with prefrontal cortex damage. *Neuropsychologia*, **32**, 881–91.

Kessler, J., Irle, E., and Markowitsch, H. J. (1986). Korsakoff and alcoholic subjects are severely impaired in animal tasks of associative memory. *Neuropsychologia*, **24**, 671–80.

Kimble, D. P. (1969). Possible inhibitory functions of the hippocampus. *Neuropsychologia*, 7, 235–44.

Kimble, D. P. (1975). Choice behavior in rats with hippocampal lesions. In *The hippocampus*, Vol. 2, *Neurophysiology and behavior* (ed. R. L. Isaacson and K. H. Pribam), pp. 309–26. Plenum Press, New York.

King, B. M., Kass, J. M., Neville, K. L., Sam, H., Tatford, A. C., III, and Zansler, C. A. (1993). Abnormal weight gain in rats with amygdaloid lesions. *Physiology and Behavior*, **54**, 467–70.

King, S. M. (1999). Escape-related behaviours in an unstable elevated and exposed environment. I. A new behavioural model of extreme anxiety. *Behavioural Brain Research*, **98**, 113–26.

King, S. M., Shehab, S., Dean, P., and Redgrave, P. (1996). Differential expression of fos-like immunoreactivity in the descending projections of superior colliculus after electrical stimulation in the rat. *Behavioural Brain Research*, **78**, 131–45.

Kinsbourne, M. and Wood, F. (1975). Short-term memory processes and the amnesic syndrome. In *Short-term memory* (ed. D. Deutsch and J. A. Deutsch), pp. 258–91. Academic Press, New York.

Kirk, I. J. and McNaughton, N. (1991). Supramammillary cell firing and hippocampal rhythmical slow activity. *NeuroReport*, **2**, 723–5.

Kirk, I. J. and McNaughton, N. (1993). Mapping the differential effects of procaine on frequency and amplitude of reticularly elicited hippocampal rhythmical slow activity. *Hippocampus*, **3**, 517–26.

Kirk, R. C., White, K. G., and McNaughton, N. (1988). Low dose scopolamine affects discriminability but not rate of forgetting in delayed conditional discrimination. *Psychopharmacology (Berlin)*, **96**, 541–6.

Kirkpatrick, B., Carter, C. S., Newman, S. W., and Insel, T. R. (1994a). Axon-sparing lesions of the medial nucleus of the amygdala decrease affiliative behaviors in the prairie vole (*Microtus ochrogaster*): behavioral and anatomical specificity. *Behavioral Neuroscience*, **108**, 501–13.

Kirkpatrick, B., Kim, J. W., and Insel, T. R. (1994b). Limbic system fos expression associated with paternal behavior. *Brain Research*, **658**, 112–18.

Kita, T., Nishijo, H., Eifuku, S., Terasawa, K., and Ono, T. (1995). Place and contingency differential responses of monkey septal neurons during conditional place–object discrimination. *The Journal of Neuroscience*, **15**, 1683–703.

Klein, D. F. (1980). Anxiety re-conceptualised. *Comprehensive Psychiatry*, **30**, 303–12.

Klein, D. F. (1981). Anxiety reconceptualized. In *Anxiety: new research and changing concepts* (ed. D. F. Klein and J. Rabkin), pp. 235–63. Raven Press, New York.

Klein, D. F. (1993). The treatment of atypical depression. *European Journal of Psychiatry*, **8**, 251–5.

Klein, D. F. (1994). Klein's suffocation theory of panic. Reply. *Archives of General Psychiatry*, **51**, 506

Klein, D. F. (1995). Recognising panic disorder. *European Journal of Psychiatry*, **10**, (Suppl. 2), 61S–63S.

Knierim, J. J., Kudrimoti, H. S., and McNaughton, B. L. (1995). Place cells, head direction cells, and the learning of landmark stability. *The Journal of Neuroscience*, **15**, 1648–59.

Koenigsberg, H. W., Pollak, C. P., Fine, J., and Kakuma, T. (1994). Cardiac and respiratory activity in panic disorder: effects of sleep and sleep lactate infusions. *American Journal of Psychiatry*, **151**, 1148–52.

Kolb, B. (1984). Functions of the frontal cortex of the rat: a comparative review. *Brain Research Reviews*, **8**, 65–98.

Kolb, B., Buhrmann, K., McDonald, R., and Sutherland, R. J. (1994). Dissociation of the medial prefrontal, posterior parietal, and posterior temporal cortex for spatial navigation and recognition memory in the rat. *Cerebral Cortex*, **6**, 664–80.

Komisaruk, B. R. (1970). Synchrony between limbic system theta activity and rhythmical behaviours in rats. *Journal of Comparative and Physiological Psychology*, **70**, 482–92.

Kondo, Y. and Arai, Y. (1995). Functional association between the medial amygdala and the medial preoptic area in regulation of mating behavior in the male rat. *Physiology and Behavior*, **57**, 69–73.

Kopchia, K. L., Altman, H. J., and Commissaris, R. L. (1992). Effects of lesions of the central nucleus of the amygdala on anxiety-like behaviors in the rat. *Pharmacology, Biochemistry and Behavior*, **43**, 453–61.

Kovacs, M., Rush, A. J., Beck, A. T., and Hollon, S. D. (1981). Depressed outpatients treated with cognitive therapy or pharmacotherapy: a one-year follow-up. *Archives of General Psychiatry*, **38**, 33–9.

Krebs, J. R., Stephens, D. W., and Sutherland, W. J. (1983). Perspectives in optimal foraging. In *Perspectives in ornithology* (ed. G. A. Clark and A. H. Brush), pp. 165–221. Cambridge University Press.

Kubie, J. L. and Ranck, J. B., Jr. (1983). Sensory-behavioural correlates in individual hippocampus neurons in three situations: space and context. In *Neurobiology of the hippocampus* (ed. W. Seifert), pp. 433–48. Academic Press, London.

La Bar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J. E., and Phelps, E. A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. *Neuron*, **20**, 937–45.

Lader, M. (1990). Introduction: use and abuse. In *In the anxiolytic jungle: where next?* (ed. D. Wheatley), pp. 3–7. Wiley, Chichester.

Lader, M. and Tyrer, P. (1975). Vegetative system and emotion. In *Emotions: their parameters and measurement* (ed. L. Levi). Raven Press, New York.

Lader, M. H. (1967). Palmar skin conductance measures in anxiety and phobic states. *Journal of Psychosomatic Research*, **11**, 271–81.

Lader, M. H. (1975). *The psychophysiology of mental illness*. Routledge and Kegan Paul, London.

Lader, M. H. and Marks, I. M. (1972). *Clinical anxiety*. Grune and Stratton, New York.

Lader, M. H. and Mathews, A. (1968). A physiological model of phobic anxiety and desensitization. *Behaviour Research and Therapy*, **6**, 411–21.

Lader, M. H. and Mathews, A. (1970). Physiological changes during spontaneous panic attacks. *Journal of Psychosomatic Research*, **14**, 377–82.

Lader, M. H. and Wing, L. (1966). *Physiological measures, sedative drugs, and morbid anxiety*, Maudsley Monographs, Vol. 18. Oxford University Press, London.

Lander, E. S. and Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185–99.

Lang, P. (1970). Stimulus control, response control, and the desensitization of fear. In *Learning approaches to therapeutic behavior* (ed. D. J. Levis), pp. 148–73. Aldine Press, Chicago.

Lavin, A. and Grace, A. A. (1994). Modulation of dorsal thalamic cell activity by the ventral pallidum: its role in the regulation of thalamocortical activity by the basal ganglia. *Synapse*, **18**, 104–27.

Leaton, R. N. and Borszcz, G. S. (1985). Potentiated startle: its relation to freezing and shock intensity in rats. *Journal of Experimental Psychology: Animal Behaviour Processes*, **11**, 421–8.

Lecrubier, Y., Puech, A. J., Azcona, A., Bailey, P. E., and Lataste, X. (1993). A randomized double-blind placebo-controlled study of tropisetron in the treatment of outpatients with generalized anxiety disorder. *Psychopharmacology (Berlin)*, **112**, 129–33.

LeDoux, J. E. (1992). Emotion and the amygdala. In *The amygdala: neurobiological aspects of emotion, memory and mental dysfunction* (ed. J. P. Aggleton), pp. 339–51. Wiley-Liss.

LeDoux, J. E. (1994). Emotion, memory and the brain. *Scientific American*, **270**, 50–9.

LeDoux, J. E. (1995). Emotion: clues from the brain. *Annual Review of Psychology*, **46**, 209–35.

LeDoux, J. E. (1996). *The emotional brain*. Simon and Schuster, New York.

LeDoux, J. E., Romanski, L., and Xagoraris, A. (1989). Indelibility of subcortical emotional memories. *Journal of Cognitive Neuroscience*, **1**, 238–43.

Lee, M. G., Chrobak, J. J., Sik, A., Wiley, R. G., and Buzsáki, G. (1994). Hippocampal theta activity following selective lesion of the septal cholinergic system. *Neuroscience*, **62**, 1033–47.

Leichnetz, G. P. and Astruc, J. (1976). The efferent projections of the medial prefrontal cortex in the squirrel monkey (*Saimiri sciureus*). *Brain Research*, **109**, 455–72.

Lelliott, P., Marks, I., McNamee, G., and Tobeña, A. (1989). Onset of panic disorder with agoraphobia. *Archives of General Psychiatry*, **46**, 1000–4.

Lenneberg, E. H. (1967). *Biological foundations of language*. Wiley, New York.

Leon, A. C., Portera, L., and Weissman, M. M. (1995). The social costs of anxiety disorders. *British Journal of Psychiatry*, **166**, (Suppl. 27), 19–22.

Leonard, C. M. (1969). The prefrontal cortex of the rat. I. Cortical projections of the medio-dorsal nucleus. II. Efferent connections. *Brain Research*, **12**, 321–43.

Levis, D. J. and Hare, N. P. (1977). A review of the theoretical rationale and empirical support for the extinction approach of implosive (flooding) therapy. In *Progressive behavior modification* (ed. M. Hersen, R. M. Eisler, and P. M. Moller), Vol. 4, pp. 299–376. Academic Press, New York.

Liebman, J. M., Mayer, D. J., and Liebeskind, J. C. (1970). Mesencephalic central gray lesions and fear-motivated behavior in rats. *Brain Research*, **23**, 353–70.

Lincoln, S., Daly, M., and Lander, E. S. (1992). Mapping genes controlling quantitative traits with MAPMaker/QTL1.1. Technical Report, Whitehead Institute, Cambridge, MA.

Lorente de No, R. (1934). Studies on the structure of the cerebral cortex. II. Continuation of the ammonic system. *Journal für Psychologie und Neurologie, Leipzig*, **46**, 113–77.

Lovell, K., Marks, I. M., Noshirvani, H., and O'Sullivan, G. (1994). Should treatment distinguish anxiogenic from anxiolytic obsessive–compulsive ruminations? Results of a pilot controlled study and of a clinical audit. *Psychotherapy and Psychosomatics*, **61**, 150–5.

Lucey, J. V., Costa, D. C., Adshead, G., Deahl, M., Busatto, G., Gacinovic, S., *et al.* (1997a). Brain blood flow in anxiety disorders—OCD, panic disorder with agoraphobia, and post-traumatic stress disorder on 99mTcHMPAO single photon emission tomography (SPET). *British Journal of Psychiatry*, **171**, 346–50.

Lucey, J. V., Costa, D. C., Busatto, G., Pilowsky, L. S., Marks, I. M., Ell, P. J., *et al.* (1997b). Caudate regional cerebral blood flow in obsessive–compulsive disorder, panic disorder and healthy controls on single photon emission computerised tomography. *Psychiatry Research: Neuroimaging Section*, **74**, 25–33.

Lucki, I., Rickels, K., Giesecke, M. A., and Geller, A. (1987). Differential effects of the anxiolytic drugs, diazepam and buspirone, on memory function. *British Journal of Clinical Pharmacology*, **23**, 207–11.

Lynch, M. A., Lindsay, J., and Ounsted, C. (1975). Tranquilizers causing aggression. *British Medical Journal*, **1** (5952), 266.

Maas, J. W. (1963). Neurochemical differences between two strains of mice. *Nature*, **197**, 255–7.

MacKay, D. G. (1987). *The organization of perception and action*. Springer, New York.

Mackintosh, N. J. (1973). Stimulus selection: learning to ignore stimuli that predict no change in reinforcement. In *Constraints on learning* (ed. R. A. Hinde and J. T. Stevenson-Hinde), pp. 75–96. Academic Press, London.

Mackintosh, N. J. (1974). *The psychology of animal learning*. Academic Press, New York.

Mackintosh, N. J. (1978). Cognitive or associative theories of conditioning: implications of an analysis of blocking. In *Cognitive processes in animal behavior* (ed. S. H. Hulse, H. Fowler, and W. K. Honig), pp. 155–75. Erlbaum, Hillsdale, NJ.

Macphail, E. M. (1996). Cognitive function in mammals: the evolutionary perspective. *Cognitive Brain Research*, **3**, 279–90.

Macrides, F., Davis, B. J., Youngs, W. M., Nadi, N. S., and Margolis, F. L. (1981). Cholinergic and catecholaminergic afferents to the olfactory bulb in the hamster: a neuroanatomical, biochemical, and histochemical investigation. *The Journal of Comparative Neurology*, **203**, 495–514.

Mahut, H. (1971). Spatial and object reversal learning in monkeys with partial temporal lobe ablations. *Neuropsychologia*, **9**, 409–24.

Maierhofer, D., Anderer, P., and Dantendorfer, K. (1998). Impaired conditional discrimination in panic disorder patients. 21st Conference of Collegium Internationale Neuropsychopharmacologicum, Glasgow, 12–16 July. (Abstract.)

Mair, W. G. P., Warrington, E. K., and Weiskrantz, L. (1979). Memory disorder in Korsakoff's psychosis: a neuropathological and neuropsychological investigation of two cases. *Brain*, **102**, 749–83.

Malsbury, C. W. and McKay, K. (1994). Neurotrophic effects of testosterone on the medial nucleus of the amygdala in adult male rats. *Journal of Neuroendocrinology*, **6**, 57–69.

Mandler, G. and Sarason, S. B. (1952). A study of anxiety and learning. *Journal of Abnormal Social Psychology*, **47**, 166–73.

Marcy, R. and Quermonne, M. A. (1974). An improved method for studying the psychogalvanic reaction in mice and its inhibition by psycholeptic drugs: comparison with the effects of other pharmacological agents. *Psychopharmacologia*, **34**, 335–49.

Markowitsch, H. J. (1982). Thalamic mediodorsal nucleus and memory: a critical evaluation of studies in animals and man. *Neuroscience and Biobehavioral Reviews*, **6**, 351–80.
Markowitsch, H. J., von Cramon, Y. D., and Schuri, U. (1993). Mnestic performance profile of a bilateral diencephalic infarct patient with preserved intelligence and severe amnesic disturbances. *Journal of Clinical and Experimental Neuropsychology*, **15**, 627–52.
Marks, I. and Marks, M. (1990). Exposure treatment of agoraphobia/panic. In *Handbook of anxiety*, Vol. 4, *The treatment of anxiety* (ed. R. Noyes, Jr., M. Roth, and G. D. Burrows), pp. 293–310. Elsevier.
Marks, I., Basoglu, M., and Noshirvani, H. (1994). Cognitive therapy in panic disorder. *British Journal of Psychiatry*, **165**, 556–7.
Marks, I. M. (1969). *Fears and phobias*. Heinemann, London.
Marks, I. M. (1973). Reduction of fear: towards a unifying theory. *Canadian Psychiatry Association Journal*, **18**, 9–12.
Marks, I. M. (1976). 'Psychopharmacology': the use of drugs combined with psychological treatment. In *Evaluation of psychological therapies* (ed. R. L. E. K. Spitzer), pp. 108–26. Johns Hopkins University Press, Baltimore.
Marks, I. M. (1987). Behavioural aspects of panic disorder. *American Journal of Psychiatry*, **144**, 1160–5.
Marks, I. M. (1988a). Overview: towards integration in panic and phobia. In *Panic and phobias*, Vol. 2 (ed. I. Hand and H. U. Wittchen), pp. 255–65. Springer, New York.
Marks, I. M. (1988b). The syndromes of anxious avoidance: classification of phobic and obsessive–compulsive phenomena. In *Handbook of anxiety*, Vol. 2, *Classification, etiological factors and associated disturbances* (ed. R. Noyes, Jr., M. Roth, and G. D. Burrows), pp. 109–46. Elsevier.
Marks, I. M. and Nesse, R. M. (1994). Fear and fitness: an evolutionary analysis of anxiety disorders. *Ethology and Sociobiology*, **15**, 247–61.
Marks, I. M., Birley, J. L. T., and Gelder, M. G. (1966). Modified leucotomy in severe agoraphobia: a controlled serial inquiry. *British Journal of Psychiatry*, **112**, 757–69.
Marks, I. M., Viswanathan, R., Lipsedge, M. S., and Gardner, R. (1972). Enhanced relief of phobias by flooding during waning diazepam effect. *British Journal of Psychiatry*, **121**, 493–505.
Marks, I. M., Swinson, R. P., Basoglu, M., Kuch, K., Noshirvani, H., O'Sullivan, G., *et al.* (1993). Alprazolam and exposure alone and combined in panic disorder with agoraphobia. *British Journal of Psychiatry*, **162**, 776–87.
Marks, J. (1990). The current role of benzodiazepines in anxiety. In *Anxiety* (ed. N. McNaughton and G. Andrews), pp. 109–13. University of Otago, Dunedin.
Markus, E. J., Barnes, C. A., McNaughton, B. L., Gladden, V. L., and Skaggs, W. E. (1994). Spatial information content and reliability of hippocampal CA1 neurons: effects of visual input. *Hippocampus*, **4**, 410–21.
Markus, E. J., Qin, Y. L., Leonard, B., Skaggs, W. E., McNaughton, B. L., and Barnes, C. A. (1995). Interactions between location and task affect the spatial and directional firing of hippocampal neurons. *The Journal of Neuroscience*, **15**, 7079–94.
Masserman, J. H. and Yum, K. S. (1946). An analysis of the influence of alcohol on experimental neuroses in cats. *Psychosomatic Medicine*, **8**, 36–52.
Matheson, G. K., Raess, B. U., and Tunnicliff, G. (1996). Effects of repeated doses of azapirones on rat brain $5HT_{1A}$ receptors and plasma corticosterone levels. *General Pharmacology*, **27**, 355–61.
Mathews, A. (1978). Fear-reduction research and clinical phobias. *Psychological Bulletin*, **85**, 390–404.

Mathews, A. (1993). Biases in processing emotional information. *The Psychologist*, **6**, 493–9.

Mathews, A. and MacLeod, C. (1985). Selective processing of threat cues in anxiety states. *Behaviour Research and Therapy*, **23**, 563–9.

Mathews, A. and MacLeod, C. (1994). Cognitive approaches to emotion and emotional disorders. *Annual Review of Psychology*, **45**, 25–50.

Mathews, A. M., Gelder, M. G., and Johnston, D. W. (1981). *Agoraphobia: nature and treatment*. Guildford Press, New York.

Matthews, G. and Gilliland, M. A. (1999). The personality theories of H. J. Eysenck and J. A Gray: a comparative review. *Personality and Individual Differences*, **26**, 583–626.

Mattick, R. P. (1990). Social phobia: an overview of psychological concepts and treatments. In *Anxiety* (ed. N. McNaughton and G. Andrews), pp. 179–85. University of Otago, Dunedin.

Maxwell, B., Powell, D. A., and Buchanan, S. L. (1994). Multiple- and single-unit activity in area 32 (prelimbic region) of the medial prefrontal cortex during Pavlovian heart rate conditioning in rabbits. *Cerebral Cortex*, **4**, 230–46.

Mayer-Gross, W., Slater, E. T. O., and Roth, M. (1979). *Clinical psychiatry*. Bailliere, Tindall, and Cox, London.

McCleary, R. A. (1966). Response-modulating functions of the limbic system: initiation and suppression. In *Progress in physiological psychology*, Vol. 1 (ed. E. Stellar and J. M. Sprague), pp. 209–72. Academic Press, New York.

McCord, P. R. and Wakefield, J. A., Jr. (1981). Arithmetic achievement as a function of introversion–extraversion and teacher-presented reward and punishment. *Personality and Individual Differences*, **2**, 145–52.

McDonald, A. J., Mascagni, F., and Guo, L. (1996). Projections of the medial and lateral prefrontal cortices to the amygdala: a *Phaseolus vulgaris* leucoagglutinin study in the rat. *Neuroscience*, **71**, 55–75.

McEwen, B. S., Gould, E. A., and Sakai, R. R. (1992). The vulnerability of the hippocampus to protective and destructive effects of glucocorticoids in relation to stress. *British Journal of Psychiatry*, **154**, 221–8.

McFarlane, A. C. (1989). The aetiology of post-traumatic morbidity: predisposing, precipitating and perpetuating factors. *British Journal of Psychiatry*, **154**, 221–8.

McGaughy, J. and Sarter, M. (1995). Behavioral vigilance in rats: task validation and effects of age, amphetamine, and benzodiazepine receptor ligands. *Psychopharmacology (Berlin)*, **117**, 340–57.

McGaughy, J., Kaiser, T., and Sarter, M. (1996). Behavioral vigilance following infusions of 192 IgG-saporin into the basal forebrain: selectivity of behavioral impairment and relation to cortical AChE-positive fiber density. *Behavioral Neuroscience*, **110**, 247–65.

McGonigle, B., McFarland, D. J., and Collier, P. (1967). Rapid extinction following drug-inhibited incidental learning. *Nature (London)*, **214**, 531–2.

McNamara, R. K. and Skelton, R. W. (1991). Diazepam impairs acquisition but not performance in the Morris water maze. *Pharmacology, Biochemistry and Behavior*, **38**, 651–8.

McNamara, R. K. and Skelton, R. W. (1992). Like diazepam, CL 218,872, a selective ligand for the benzodiazepine Omega-1 receptor subtype, impairs place learning in the Morris water maze. *Psychopharmacology (Berlin)*, **107**, 347–51.

McNamara, R. K. and Skelton, R. W. (1993). Effects of intracranial infusions of chlordiazepoxide on spatial learning in the Morris water maze. I. Neuroanatomical specificity. *Behavioural Brain Research*, **59**, 175–91.

McNaughton, B. L., Leonard, B., and Chen, L. (1989). Cortical–hippocampal interactions and cognitive mapping: a hypothesis based on reintegration of the parietal and inferotemporal pathways for visual processing. *Psychobiology*, **17**, 236–46.

McNaughton, B. L., Chen, L. L., and Markus, E. J. (1991). 'Dead reckoning,' landmark learning, and the sense of direction: a neurophysiological and computational hypothesis. *Journal of Cognitive Neuroscience*, **3**, 190–202.

McNaughton, B. L., Barnes, C. A., Gerrard, J. L., Gothard, K., Jung, M. W., Knierim, J. J., *et al.* (1996). Deciphering the hippocampal polyglot: the hippocampus as a path integration system. *The Journal of Experimental Biology*, **199**, 173–85.

McNaughton, B. L., Knierim, J. J., and Wilson, M. A. (1997). Vector coding and the vestibular foundations of spatial cognition: neurophysiological and computational mechanisms. In *The cognitive neurosciences* (ed. M. S. Gazzaniga), pp. 585–95. MIT Press, Cambridge.

McNaughton, N. (1985a). Is the hippocampus a store, intermediate or otherwise? *Behavioural and Brain Sciences*, **8**, 508–9.

McNaughton, N. (1985b). The effects of systemic and intraseptal injections of sodium amylobarbitone on rearing and ambulation in rats. *Australian Journal of Psychology*, **37**, 15–27.

McNaughton, N. (1985c). Chlordiazepoxide and successive discrimination: different effects on acquisition and performance. *Pharmacology, Biochemistry and Behavior*, **23**, 487–94.

McNaughton, N. (1989a). *Biology and emotion*. Cambridge University Press.

McNaughton, N. (1989b). Anxiety: one label for many processes. *New Zealand Journal of* Psychology, **18**, 51–9.

McNaughton, N. (1993). Stress and behavioural inhibition. In *Stress: an integrated approach* (ed. S. C. Stanford and P. Salmon), pp. 191–206. Academic Press.

McNaughton, N. (1994). The hippocampus: relational processor or anti-processor? *Behavioural and Brain Sciences*, **17**, 487–8.

McNaughton, N. (1997). Cognitive dysfunction resulting from hippocampal hyperactivity—a possible cause of anxiety disorder. *Pharmacology, Biochemistry and Behavior*, **56**, 603–11.

McNaughton, N. and Coop, C. F. (1991). Neurochemically dissimilar anxiolytic drugs have common effects on hippocampal rhythmic slow activity. *Neuropharmacology*, **30**, 855–63.

McNaughton, N. and Mason, S. T. (1980). The neuropsychology and neuropharmacology of the dorsal ascending noradrenergic bundle—a review. *Progress in Neurobiology*, **14**, 157–219.

McNaughton, N. and Morris, R. G. M. (1987). Chlordiazepoxide, an anxiolytic benzodiazepine, impairs place navigation in rats. *Behavioural Brain Research*, **24**, 39–46.

McNaughton, N. and Morris, R. G. M. (1992). Buspirone produces a dose-related impairment in spatial navigation. *Pharmacology, Biochemistry and Behavior*, **43**, 167–71.

McNaughton, N. and Sedgwick, E. M. (1978). Reticular stimulation and hippocampal theta rhythm in rats: effects of drugs. *Neuroscience*, **2**, 629–32.

McNaughton, N., James, D. T. D., Stewart, J., Gray, J. A., Valero, I., and Drewnowski, A. (1977). Septal driving of hippocampal theta rhythm as a function of frequency in the male rat: effects of drugs. *Neuroscience*, **2**, 1019–27.

McNaughton, N., Azmitia, E. C., Williams, J. H., Buchan, A., and Gray, J. A. (1980a). Septal elicitation of hippocampal theta rhythm after localized de-afferentation of serotoninergic fibres. *Brain Research*, **200**, 259–69.

McNaughton, N., Davis, N. M., Brookes, S., Rawlins, J. N. P., Feldon, J., and Gray, J. A. (1980b). Minor tranquillisers and hippocampal connections—some behavioural dissociations. In *Psychophysiology* (ed. R. Sinz and M. R. Rosenzweig), pp. 183–9. Elsevier Biomedical Press, Amsterdam.

McNaughton, N., Owen, S., Boarder, M. R., Gray, J. A., and Fillenz, M. (1984). Responses to novelty in rats with lesions of the dorsal noradrenergic bundle. *New Zealand Journal of Psychology*, **13**, 16–24.

McNaughton, N., Richardson, J., and Gore, C. (1986). Reticular elicitation of hippocampal slow waves: common effects of some anxiolytic drugs. *Neuroscience*, **19**, 899–903.

McNaughton, N., Logan, B., Panickar, K. S., Kirk, I. J., Pan, W. X., Brown, N. T., *et al.* (1995). Contribution of synapses in the medial supramammillary nucleus to the frequency of hippocampal theta rhythm in freely moving rats. *Hippocampus*, **5**, 534–45.

McNaughton, N., Panickar, K. S., and Logan, B. (1996). The pituitary–adrenal axis and the different behavioral effects of buspirone and chlordiazepoxide. *Pharmacology, Biochemistry and Behavior*, **54**, 51–6.

McNish, K. A., Gewirtz, J. C., and Davis, M. (1997). Evidence of contextual fear after lesions of the hippocampus: a disruption of freezing but not fear-potentiated startle. *The Journal of Neuroscience*, **17**, 9353–60.

Meck, W. H. (1988). Hippocampal function is required for feedback control of an internal clock's criterion. *Behavioral Neuroscience*, **102**, 54–60.

Meck, W. H., Church, R. M., Wenk, G. L., and Olton, D. S. (1987). Nucleus basalis magnocellularis and medial septal area lesions differentially impair temporal memory. *The Journal of Neuroscience*, 7, 3505–11.

Medina, J. H., Paladini, A. C., and Izquierdo, I. (1993). Naturally occurring benzodiazepines and benzodiazepine-like molecules in brain. *Behavioural Brain Research*, **58**, 1–8.

Meibach, R. C. and Siegel, A. (1977). Subicular projections to the posterior cingulate cortex in rats. *Experimental Neurology*, **57**, 264–74.

Meldrum, B. S. (1975). Epilepsy and gamma-aminobutyric acid-mediated inhibition. *International Review of Neurobiology*, **17**, 1–36.

Meunier, M., Bachevalier, J., Mishkin, M., and Murray, E. A. (1993). Effects on visual recognition of combined and separate ablations of the entorhinal and perirhinal cortex in rhesus monkeys. *The Journal of Neuroscience*, **13**, 5418–32.

Meunier, M., Hadfield, W., Bachevalier, J., and Murray, E. A. (1996). Effects of rhinal cortex lesions combined with hippocampectomy on visual recognition memory in rhesus monkeys. *Journal of Neurophysiology*, **75**, 1190–205.

Miczek, K. A. (1973). Effects of scopolamine, amphetamine and benzodiazepines on conditioned suppression. *Pharmacology, Biochemistry and Behavior*, **1**, 401–11.

Miczek, K. A., Weerts, E. M., Vivian, J. A., and Barros, H. M. (1995). Aggression, anxiety and vocalizations in animals: $GABA_A$ and 5HT anxiolytics. *Psychopharmacology (Berlin)*, **121**, 38–56.

Millenson, J. R. (1967). *Principles of behavioral analysis*. Macmillan, New York.

Miller, C. L., Bickford, P. C., Wiser, A. K., and Rose, G. M. (1995). Long-term potentiation disrupts auditory gating in the rat hippocampus. *The Journal of Neuroscience*, **15**, 5820–30.

Miller, G. A., Galanter, E. H., and Pribram, K. H. (1960). *Plans and the structure of behavior*. Rinehart and Winston, New York.

Miller, N. E. (1951). Learnable drives and rewards. In *Handbook of experimental psychology* (ed. S. S. Stevens), pp. 435–72. Wiley, New York.

Miller, R. (1991). *Cortico-hippocampal interplay and the representation of contexts in the brain*. Springer, Berlin.
Miller, R. (1995). *Cortico-hippocampal interplay: self-organizing phase-locked loops for indexing memory*. Unpublished manuscript.
Miller, V. M. and Best, P. J. (1980). Spatial correlates of hippocampal unit activity are altered by lesions of the fornix and entorhinal cortex. *Brain Research*, **194**, 311–23.
Milner, B. (1963). Effects of different brain lesions on card sorting. *Archives of Neurology*, **9**, 90–100.
Milner, B. (1964). Some effects of frontal lobectomy in man. In *The frontal granular cortex and behavior* (ed. J. M. Warren and K. Akert), pp. 313–31. McGraw-Hill, New York.
Milner, B. (1968). Preface: material specific and generalized memory loss. *Neurospychologia*, **6**, 175–9.
Milner, B. (1970). Memory and the medial temporal regions of the brain. In *Biology of memory* (ed. K. H. Pribram and D. E. Broadbent), pp. 29–50. Academic Press, New York.
Milner, B. (1971). Interhemispheric differences and psychological processes. *British Medical Bulletin*, **27**, 272–7.
Milner, B. (1995). Aspects of human frontal lobe function. *Advances in Neurology*, **66**, 67–84.
Milner, B., Corkin, S., and Teuber, H.-L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of H.M. *Neuropsychologia*, **6**, 215–34.
Mineka, S. (1987). A primate model of phobic fears. In *Theoretical foundations of behavior therapy* (ed. H. J. Eysenck and I. Martin), pp. 81–111. Plenum, New York.
Mineka, S. (1992). Evolutionary memories, emotional processing, and the emotional disorders. *The Psychology of Learning and Motivation*, **28**, 161–206.
Mineka, S. and Sutton, S. K. (1992). Cognitive biases and the emotional disorders. *Psychological Science*, **3**, 65–9.
Mineka, S. and Zinbarg, R. (1995). Conditioning and ethological models of social phobia. In *Social phobia: diagnosis, assessment and treatment* (ed. R. Heimberg, M. Liebowitz, D. Hope, and F. Schneier), pp. 134–62. Guildford Press.
Mineka, S. and Zinbarg, R. (1996). Conditioning and ethological models of anxiety disorders: stress-in-dynamic-context anxiety models. In *Perspectives on anxiety, panic and fear*, Vol. 43, *Nebraska Symposium on motivation* (ed. D. Hope), pp. 135–210. University of Nebraska Press, Lincoln.
Minerbo, G., Albeck, D., Goldberg, E., Lindberg, T., Nakari, M., Martinez, C., *et al.* (1994). Activity of peptidergic neurons in the amygdala during sexual behavior in the male rat. *Experimental Brain Research*, **97**, 444–50.
Miyamoto, Y. and Jinnai, K. (1994). The inhibitory input from the substantia nigra to the mediodorsal nucleus neurons projecting to the prefrontal cortex in the cat. *Brain Research*, **649**, 313–18.
Mizumori, S. J. Y. and Williams, J. D. (1993). Directionally-selective mnemonic properties of neurons in the lateral dorsal nucleus of the thalamus of rats. *The Journal of Neuroscience*, **13**, 4015–28.
Mizumori, S. J. Y., Miya, D. Y., and Ward, K. E. (1994). Reversible inactivation of the lateral dorsal thalamus disrupts hippocampal place representation and impairs spatial learning. *Brain Research*, **644**, 168–74.
Modigh, K. (1990). The pharmacology of clomipramine in anxiety disorders. In *Anxiety* (ed. N. McNaughton and G. Andrews), pp. 93–100. Otago University Press, Dunedin.

Mogenson, G. J. and Nielsen, M. (1984). A study of the contribution of hippocampal-accumbens-subpallidal projections to locomotor activity. *Behavioral and Neural Biology*, **42**, 52–60.

Möhler, H. and Okada, T. (1977). Benzodiazepine receptor: demonstration in the central nervous system. *Science (NY)*, **198**, 849–51.

Möhler, H., Fritschy, J. M., and Randolph, U. (2002). A new benzodiazepine pharmacology. *Journal of Pharmacology and Experimental Therapeutics*, 300, 2–8.

Money, E. A., Kirk, R. C., and McNaughton, N. (1992). Alzheimer's dementia produces a loss of discrimination but no increase in rate of memory decay in delayed matching to sample. *Neuropsychologia*, **30**, 133–45.

Moore, H., Sarter, M., and Bruno, J. P. (1995). Bidirectional modulation of cortical acetylcholine efflux by infusion of benzodiazepine receptor ligands into the basal forebrain. *Neuroscience Letters*, **189**, 31–4.

Morecraft, R. J. and Van Hoesen, G. W. (1998). Convergence of limbic input to the cingulate motor cortex in the rhesus monkey. *Brain Research Bulletin*, **45**, 209–32.

Morris, J., Firth, C. D., Perrett, D. I., Rowland, D. I., Young, A. W., Calder, A. J., *et al.* (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, **383**, 812–15.

Morris, R. G., Pickering, A., Abrahams, S., and Feigenbaum, J. D. (1996). Space and the hippocampal formation in humans. *Brain Research Bulletin*, **40**, 487–90.

Morris, R. G. M. (1981). Spatial localisation does not require the presence of local cues. *Learning and Motivation*, **12**, 239–60.

Morris, R. G. M. (1984). Development of a water-maze procedure for the study of spatial learning in the rat. *Journal of Neuroscience Methods*, **11**, 47–60.

Morris, R. G. M., Garrud, P., Rawlins, J. N. P., and O'Keefe, J. (1982). Place navigation in rats with hippocampal lesions. *Nature (London)*, **297**, 681–3.

Morris, R. G. M., Schenk, F., Tweedie, F., and Jarrard, L. E. (1990). Ibotenate lesions of hippocampus and/or subiculum: dissociating components of allocentric spatial learning. *European Journal of Neuroscience*, **2**, 1016–28.

Mowrer, O. H. (1960). *Learning theory and behavior*. Wiley, New York.

Muller, R. U. and Kubie, J. L. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *The Journal of Neuroscience*, 7, 1951–68.

Muller, R. U., Kubie, J. L., and Ranck, J. B., Jr. (1987). Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. *The Journal of Neuroscience*, 7, 1935–50.

Mumby, D. G. (1995). Assessing working memory for objects in rats: no one said it was easy. *NeuroReport*, **6**, 1960–2.

Murphy, G. E., Simons, A. D., Wetzel, R. D., and Lustman, P. J. (1984). Cognitive therapy and pharmacotherapy: singly and together in the treatment of depressed outpatients. *Archives of General Psychiatry*, **41**, 33–41.

Murray, E. A. (1996). What have ablation studies told us about the neural substrates of stimulus memory? *The Neurosciences*, **8**, 13–22.

Myers, J. S. (1971). Some effects of noncontingent aversive stimulation. In *Aversive conditioning and learning* (ed. F. R. Brush), pp. 469–536. Academic Press, New York.

Nadel, L. (1991). The hippocampus and space revisited. *Hippocampus*, **1**, 221–9.

Nadel, L. and Jacobs, W. J. (1996). The role of the hippocampus in PTSD, panic and phobia. In *The hippocampus: functions and clinical relevance* (ed. N. Kato), pp. 455–63. Elsevier.

Näring, G. W. B. and Van der Staak, C. P. F. (1995). Perception of heart rate and blood pressure: the role of alexithymia and anxiety. *Psychotherapy and Psychosomatics*, **63**, 193–200.

Nelson, E., Early, T. S., and Haller, J. W. (1993). Visual attention in obsessive–compulsive disorder. *Psychiatry Research*, **49**, 183–96.

Nesse, R. M. and Klaas, R. (1994). Risk perception by patients with anxiety disorders. *Journal of Nervous and Mental Diseases*, **182**, 465–70.

Newman, J. P. (1987). Reaction to punishment in extroverts and psychopaths: implications for the impulsive behaviour of disinhibited individuals. *Journal of Research in Personality*, **21**, 464–80.

Nielsen, M., Braestrup, C., and Squires, R. F. (1978). Evidence for a late evolutionary appearance of brain-specific benzodiazepine receptors: an investigation of 18 vertebrate and 5 invertebrate species. *Brain Research*, **141**, 342–6.

Noyes, R. Jr., Kathol, R., Clancy, J., and Crowe, R. R. (1981). Antianxiety effects of propranolol: a review of clinical studies. In *Anxiety*: *new research and changing concepts* (ed. D. F. Klein and J. Rabkin), pp. 81–93. Raven Press, New York.

Noyes, R., Jr., Roth, M., and Burrows, G. D. (1988). *Handbook of Anxiety, Vol. 2: Classification, etiological factors and associated disturbances*. Elsevier, Amsterdam.

Numan, M. (1994). A neural circuitry analysis of maternal behavior in the rat. *Acta Paediatrica*, **83**, (Suppl. 397), 19–28.

Nutt, D. J. (1996). The pharmacology of human anxiety. *Pharmacology and Therapeutics*, **47**, 233–66.

Oakley-Browne, M. A., Frampton, C. M., Joyce, P., Wells, J. E., Bushnell, J. A., and Hornblow, A. R. (1990). The epidemiology of generalised anxiety disorder: data from a New Zealand community. In *Anxiety* (ed. N. McNaughton and G. Andrews), pp. 40–52. University of Otago, Dunedin.

Ogden, J. and Corkin, S. (1991). Memories of H.M. In *Memory mechanisms: a tribute to G.V. Goddard* (ed. W. C. Abraham, M. Corballis, and K. G. White), pp. 195–215. Erlbaum, Hillsdale, NJ.

O'Hanlon, J. F., Vermeeren, A., Uiterwijk, M. M. C., Van Veggel, L. M. A., and Swijgman, H. F. (1995). Anxiolytics' effects on the actual driving performance of patients and healthy volunteers in a standardized test. An integration of three studies. *Neuropsychobiology*, **31**, 81–8.

Öhman, A. (1971). Differentiation of conditioned and orienting response components in electrodermal conditioning. *Psychophysiology*, **8**, 7–22.

Öhman, A. (1979). Fear relevance, autonomic conditioning, and phobias: a laboratory model. In *Trends in behavior therapy* (ed. P. O. Sjoden, S. Bates, and W. W. Dockens), pp. 107–33. Academic Press, New York.

Öhman, A. and Dimberg, U. (1978). Facial expressions as conditioned stimuli for electrodermal responses: a case of 'preparedness'? *Journal of Personality and Social Psychology*, **36**, 1251–8.

Öhman, A. and Soares, J. J. F. (1998). Emotional conditioning to masked stimuli: expectancies for aversive outcomes following nonrecognized fear-relevant stimuli. *Journal of Experimental Psychology: General*, **127**, 69–82.

Öhman, A., Eriksson, A., Fredrikson, M., Hugdahl, K., and Olofsson, C. (1974). Habituation of the electrodermal orienting reaction to potentially phobic and supposedly neutral stimuli in normal human subjects. *Biological Psychology*, **2**, 85–93.

Öhman, A., Eriksson, A., and Olofsson, C. (1975). One-trial learning and superior resistance to extinction of autonomic responses conditioned to potentially phobic stimuli. *Journal of Comparative and Physiological Psychology*, **88**, 619–27.

Öhman, A., Fredikson, M., Hugdahl, K., and Rimmo, P.-A. (1976). The premise of equipotentiality in human classical conditioning: conditioned electrodermal responses to potentially phobic stimuli. *Journal of Experimental Psychology: General*, **105**, 313–37.

Öhman, A., Fredrikson, M., and Hugdahl, K. (1978). Orienting and defensive responding in the electrodermal system: palmar–dorsal differences and recovery rate during conditioning to potentially phobic stimuli. *Psychophysiology*, **15**, 93–101.

Okaichi, Y. and Okaichi, H. (1994). Effects of fimbria–fornix lesions on avoidance tasks with temporal elements in rats. *Physiology and Behavior*, **56**, 759–65.

Okasha, A., Bishry, Z., Khalil, A. H., Darwish, T. A., Seif El Dawla, A., and Shohdy, A. (1994). Panic disorder. An overlapping or independent entity. *British Journal of Psychiatry*, **164**, 818–25.

O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Experimental Neurology*, **51**, 78–109.

O'Keefe, J. and Black, A. H. (1978). Single unit and lesion experiments on the sensory inputs to the hippocampal map. In *Functions of the septo-hippocampal system*, Ciba Foundation Symposium 58 (New Series) (ed. K. Elliott and J. Whelan), pp. 179–92. Elsevier, Amsterdam.

O'Keefe, J. and Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. *Nature*, **381**, 425–8.

O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, **34**, 171–5.

O'Keefe, J. and Nadel, L. (1978). *The hippocampus as a cognitive map*. Clarendon Press, Oxford.

O'Keefe, J. and Speakman, A. (1987). Single unit activity in the rat hippocampus during a spatial memory task. *Experimental Brain Research*, **68**, 1–27.

Olsen, R. W., Greenlee, D., Van Ness, P., and Ticku, M. K. (1978a). Studies on the gamma-aminobutyric acid receptor/ionophore proteins in mammalian brain. In *Amino acids as chemical transmitters* (ed. F. Fonnum), pp. 467–86. Plenum Press, New York.

Olsen, R. W., Ticku, M. K., Greenlee, D., and Van Ness, P. (1978b). GABA receptor and ionophore binding sites: interaction with various drugs. In *GABA-neurotransmitters* (ed. P. Krogsgaard-Larsen, J. Scheel-Kruger, and H. Kofod), pp. 165–78. Academic Press, New York.

Olton, D. S. (1978a). Characteristics of spatial memory. In *Cognitive processes in animal behavior* (ed. S. H. Hulse, H. Fowler, and W. K. Honig), pp. 341–73. Erlbaum, Hillsdale, NJ.

Olton, D. S. (1978b). The function of septo-hippocampal connections in spatially organized behaviour. In *Functions of the septo-hippocampal system*, Ciba Foundation Symposium 58 (New Series) (ed. K. Elliott and J. Whelan), pp. 327–42. Elsevier, Amsterdam.

Olton, D. S. and Isaacson, R. L. (1968). Hippocampal lesions and active avoidance. *Physiology and Behavior*, **3**, 719–24.

Olton, D. S. and Papas, B. C. (1979). Spatial memory and hippocampal function. *Neuropsychologia*, **17**, 669–82.

Olton, D. S. and Werz, M. A. (1978). Hippocampal function and behaviour: spatial discrimination and response inhibition. *Physiology and Behavior*, **20**, 597–606.

Olton, D. S., Branch, M., and Best, P. J. (1978a). Spatial correlates of hippocampal unit activity. *Experimental Neurology*, **58**, 387–409.

Olton, D. S., Walker, J. A., and Gage, F. H. (1978b). Hippocampal connections and spatial discrimination. *Brain Research*, **139**, 295–308.

Olton, D. S., Becker, J. T., and Handelmann, G. E. (1979a). Hippocampus, space and memory. *Behavioural and Brain Sciences*, **2**, 313–22.

Olton, D. S., Becker, J. T., and Handelmann, G. E. (1979b). Hippocampus, space, and memory: authors' responses. *Behavioural and Brain Sciences*, **2**, 352–9.

Olton, D. S., Wible, C. G., and Shapiro, M. L. (1986). Mnemonic theories of hippocampal function. *Behavioral Neuroscience*, **100**, 852–5.

O'Mara, S. M. (1995). Spatially selective firing properties of hippocampal formation neurons in rodents and primates. *Progress in Neurobiology*, **45**, 253–74.

O'Mara, S. M., Rolls, E. T., Berthoz, A., and Kesner, R. P. (1994). Neurons responding to whole-body motion in the primate hippocampus. *The Journal of Neuroscience*, **14**, 6511–23.

Oscar-Berman, M. O. and Samuels, I. (1973). Stimulus-preference and memory factors in Korsakoff's syndrome. Paper presented at the American Psychological Association Meeting, Montreal.

Oscar-Berman, M. O. and Samuels, I. (1977). Stimulus preference and memory factors in Korsakoff's syndrome. *Neuropsychologia*, **15**, 99–106.

Page, A. C. (1994). Distinguishing panic disorder and agoraphobia from social phobia. *Journal of Nervous and Mental Diseases*, **182**, 611–17.

Pan, W. X. and McNaughton, N. (1997). The medial supramammillary nucleus, spatial learning and the frequency of hippocampal theta activity. *Brain Research*, **764**, 101–8.

Pandya, D., Dye, P., and Butters, N. (1971). Efferent cortico-cortical projections of the prefrontal cortex in the Rhesus monkey. *Brain Research*, **31**, 35–46.

Panksepp, J. (1982). Towards a general psychobiological theory of emotions. *Behavioral and Brain Sciences*, **5**, 407–68.

Papez, J. W. (1937). A proposed mechanism of emotion. *Archives of Neurological Psychiatry*, **38**, 725–43.

Papp, L. A., Klein, D. F., and Gorman, J. M. (1993). Carbon dioxide hypersensitivity, hyperventilation, and panic disorder. *American Journal of Psychiatry*, **150**, 1149–57.

Papp, L. A., Martinez, J. M., Klein, D. F., Coplan, J. D., and Gorman, J. M. (1995). Rebreathing tests in panic disorder. *Biological Psychiatry*, **38**, 240–5.

Parkin, A. J., Rees, J. E., Hunkin, N. M., and Rose, P. E. (1994). Impairment of memory following discrete thalamic infarction. *Neuropsychologia*, **32**, 39–51.

Parmeggiani, P. L., Azzaroni, A., and Lenzi, P. (1971). On the functional significance of the circuit of Papez. *Brain Research*, **30**, 357–74.

Parmeggiani, P. L., Lenzi, P. L., and Azzaroni, A. (1974). Transfer of the hippocampal output by the anterior thalamic nuclei. *Brain Research*, **67**, 269–78.

Passingham, R. E. (1993). *The frontal lobes and voluntary action*. Oxford University Press.

Pavlov, I. P. (1927). *Conditioned reflexes* (trans. G. V. Anrep). Oxford University Press, London.

Pellow, S. and File, S. E. (1986). Anxiolytic and anxiogenic drug effects on exploratory activity in an elevated plus-maze: a novel test of anxiety in the rat. *Pharmacology, Biochemistry and Behavior*, **24**, 525–9.

Pesold, C. and Treit, D. (1994). The septum and amygdala differentially mediate the anxiolytic effects of benzodiazepines. *Brain Research*, **638**, 295–301.

Pesold, C. and Treit, D. (1995). The central and basolateral amygdala differentially mediate the anxiolytic effects of benzodiazepines. *Brain Research*, **671**, 213–21.

Petrides, M. (1994). Frontal lobes and working memory: evidence from investigations of the effects of cortical excisions in nonhuman primates. In *Handbook of neuropsychology* (ed. F. Boller and J. Grafman), pp. 59–82. Elsevier, Montreal.

Petrides, M. and Pandya, D. N. (1994). Comparative architectonic analysis of the human and the macaque frontal cortex. In *Handbook of neuropsychology* (ed. F. Boller and J. Grafman), pp. 17–58. Elsevier, Montreal.

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., *et al.* (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, **389**, 495–8.

Phillips, M. L., Young, A. W., Scott, S. K., Calder, A. J., Andrew, C., Giampietro, V., *et al.* (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proceedings of the Royal Society of London B: Biological Sciences*, **265**, 1809–817.

Phillips, R. G. and LeDoux, J. E. (1994). Lesions of the dorsal hippocampal formation interfere with background but not foreground contextual fear conditioning. *Learning and Memory*, **1**, 34–44.

Pickering, A. D. (1999). Personality correlates of the dopamine facilitation of incentive motivation: impulsive sensation seeking rather than extraversion? *Behavioural and Brain Sciences*, **22**, 534–5.

Pickering, A. D. and Gay, J. A. (1999). The neuroscience of personality. In *Handbook of personality theory and research (Second edition)* (ed. L. A. Perrin and O.P. John), pp. 277–99. Guilford Press, New York, London.

Pickering, A. D., Corr, P. J., Powell, J. H., Kumari, V., Thornton, J. C., and Gray, J. A. (1997). Individual differences in reactions to reinforcing stimuli are neither black nor white: to what extent are they gray? In *The scientific study of human nature: tribute to Hans J. Eysenck at eighty* (ed. H. Nyborg), pp. 36–67. Elsevier, London.

Pickering, A. D., Corr, P. J., and Gray, J. A. (1998). Interactions and reinforcement sensitivity theory: a theoretical analysis of Rusting and Larsen. *Personality and Individual Differences*, **26**, 357–65.

Pigott, S. and Milner, B. (1994). Capacity of visual short-term memory after unilateral frontal or anterior temporal-lobe resection. *Neuropsychologia*, **32**, 969–81.

Pinel, J. P. J. (1997). *Biopsychology*. Allyn and Bacon, Boston.

Pitkänen, A., Savander, V., and LeDoux, J. E. (1997). Organization of intra-amygdaloid circuitries in the rat: an emerging framework for understanding functions of the amygdala. *Trends in Neuroscience*, **20**, 517–23.

Pitman, R. K. (1989). Posttraumatic stress disorder hormones, and memory [editorial]. *Biological Psychiatry*, **26**, 221–3.

Pitman, R. K., Orr, S. P., and Shalev, A. Y. (1993). Once bitten, twice shy: beyond the conditioning model of PTSD. *Biological Psychiatry*, **33**, 145–6.

Pittendrigh, C. S. (1958). Adaptation, natural selection and behaviour. In *Behaviour and evolution* (ed. A. Roes and G. G. Simpson). Yale University Press.

Ploghaus, A., Narain, C., Beckmann, C. F., Clare, S., Bantick, S., Wise, R., Matthews, P. M., Rawlins, S. N., and Tracey, I. (2001). Exacerbation of pain by anxiety is associated with activity in a hippocampal network. *Journal of Neuroscience*, 21, 9896–903.

Plomin, R. P., Owen, M. J., and McGuffin, P. (1994). The genetic basis of complex human behaviours. *Science*, **264**, 1733–79.

Plutchik, R. (1962). *The emotions: facts, theories and a new model*. Random House, New York.

Poucet, B., Thinus-Blanc, C., and Muller, R. U. (1994). Place cells in the ventral hippocampus of rats. *NeuroReport*, **5**, 2045–8.

Powell, G. E. (1979). *Brain and personality*. Saxon House, London.

Power, K. G., Simpson, R. J., Swanson, V., Wallace, L. A., Feistner, A. T. C., and Sharp, D. (1990). A controlled comparison of cognitive-behaviour therapy, diazepam and placebo, alone and in combination for the treatment of generalised anxiety disorder. *Journal of Anxiety Disorders*, **4**, 267–92.

Proctor, S. (1969). Duration of exposure to items and pretreatment training as factors in systematic desensitization therapy. In *Advances in behaviour therapy 1968* (ed. R. D. Rubin and C. M. Franks), pp. 105–16. Academic Press, New York.

Quirk, G. J., Muller, R. U., and Kubie, J. L. (1990). The firing of hippocampal place cells in the dark depends on the rat's recent experience. *The Journal of Neuroscience*, **10**, 2008–17.

Quirk, G. J., Muller, R. U., Kubie, J. L., and Ranck, J. B. J. (1992). The positional firing properties of medial entorhinal neurons: description and comparison with hippocampal place cells. *The Journal of Neuroscience*, **12**, 1945–63.

Rachman, S. (1978). An anatomy of obsessions. *Behaviour Analysis and Modification*, **2**, 253–78.

Rachman, S. and Hodgson, R. (1979). *Obsessions and compulsions*. Prentice Hall, New York.

Rachman, S. and Seligman, M. E. P. (1976). Unprepared phobias: be prepared. *Behaviour Research and Therapy*, **14**, 333–8.

Rajkowska, G. and Goldman-Rakic, P. S. (1995a). Cytoarchitectonic definition of prefrontal areas in the normal human cortex. I. Remapping of areas 9 and 46 using quantitative criteria. *Cerebral Cortex*, **5**, 307–22.

Rajkowska, G. and Goldman-Rakic, P. S. (1995b). Cytoarchitectonic definition of prefrontal areas in the normal human cortex. II. Variability in locations of areas 9 and 46 and relationship to the Talairach coordinate system. *Cerebral Cortex*, **5**, 323–37.

Ramon y Cajal, S. (1955). *Histologie du système nerveux de l'homme et des vertébrates, Vol. II (1911)*. Consejo Superior de Investigaciones Cientificos Instituto Ramon y Cajal, Madrid.

Ranck, J. B., Jr. (1973). Studies on single neurons in dorsal hippocampal formation and septum in unrestrained rats. *Experimental Neurology*, **41**, 461–555.

Randall, L. O. and Kappell, B. (1973). Pharmacological activity of some benzodiazepines and their metabolites. In *The benzodiazepines* (ed. S. Garattini, E. Mussini, and L. O. Randall), pp. 27–51. Raven Press, New York.

Ransom, B. R. and Barker, J. L. (1976). Pentobarbital selectively enhances GABA-mediated post-synaptic inhibition in tissue cultured mouse spinal neurons. *Brain Research*, **114**, 530–5.

Raphelson, A. C., Isaacson, R. L., and Douglas, R. J. (1965). The effect of distracting stimuli on the runway performance of limbic damaged rats. *Psychonomic Science*, **3**, 483–4.

Rapoport, J. L. (1989). The biology of obsessions and compulsions. *Scientific American*, 63–9.

Rauch, S. L., Savage, C. R., Baer, L., Breiter, H. C., Alpert, N. M., Fischman, A. J., *et al.* (1995). Reciprocal lateralization of frontal lobe functions. Reply. *Archives of General Psychiatry*, **52**, 159–60.

Rauch, S. L., Van der Kolk, B. A., Fisler, R. E., Alpert, N. M., Orr, S. P., Savage, C. R., *et al.* (1996). A symptom provocation study of posttraumatic stress disorder using positron emission tomography and script-driven imagery. *Archives of General Psychiatry*, **53**, 380–7.

Rawlins, J. N. P. (1977). Behavioural and physiological correlates of limbic system activity. Unpublished D.Phil. thesis, University of Oxford.

Rawlins, J. N. P. (1985). Association across time: the hippocampus as a temporary memory store. *Behavioural and Brain Sciences*, **8**, 479–96.

Rawlins, J. N. P. and Tsaltas, E. (1983). The hippocampus, time and working memory. *Behavioural Brain Research*, **10**, 233–62.

Rawlins, J. N. P., Lyford, G. L., Seferiades, A., Deacon, R. M. J., and Cassaday, H. J. (1993). Critical determinants of nonspatial working memory deficits in rats with conventional lesions of the hippocampus or fornix. *Behavioral Neuroscience*, **107**, 420–33.

Razran, G. (1971). *Mind in evolution*. Houghton Mifflin, Boston.

Redgrave, P. and Dean, P. (1997). Does the PAG learn about emergencies from the superior colliculus? In *The midbrain periaqueductal gray matter* (ed. A. Depaulis and R. Bandler), pp. 199–209. Plenum Press, New York.

Rempel-Clower, N. L., Zola, S. M., Squire, L. R., and Amaral, D. G. (1996). Three cases of enduring memory impairment after bilateral damage limited to the hippocampal formation. *The Journal of Neuroscience*, **16**, 5233–55.

Rescorla, R. A. (1978). Some implications of a cognitive perspective on Pavlovian conditioning. In *Cognitive processes in animal behavior* (ed. S. H. Hulse, H. Fowler, and W. K. Honig), pp. 15–50. Wiley, New York.

Revusky, S. (1971). The role of interference in association over a delay. In *Animal memory* (ed. W. K. Honig and P. H. R. James), pp. 155–213. Academic Press, New York.

Reynolds, D. S., McKernan, R. M., and Dawson, G. R. (2001). Anxiolytic-like action of diazepam: which GABA(A) receptor subtype is involved? *Trends in Pharmacological Sciences*, **22**, 402–3.

Richards, G., Schoch, P., and Haefely, W. (1991). Benzodiazepine receptors: new vistas. *The Neurosciences*, **3**, 191–203.

Rickels, K. (1978). Use of anti-anxiety agents in anxious outpatients. *Psychopharmacology (Berlin)*, **58**, 1–17.

Rickert, E. J., Bennett, T. L., Lane, P., and French, J. (1978). Hippocampectomy and the attenuation of blocking. *Behavioural Biology*, **22**, 147–60.

Ridley, R. M., Timothy, C. J., MacLean, C. J., and Baker, H. F. (1995). Conditional learning and memory impairments following neurotoxic lesion of the CA1 field of the hippocampus. *Neuroscience*, **67**, 263–75.

Ringo, J. L. (1988). Seemingly discrepant data from hippocampectomized macaques are reconciled by detectability analysis. *Behavioral Neuroscience*, **102**, 173–7.

Ringo, J. L. (1991). Memory decays at the same rate in macaques with and without brain lesions when expressed in d' or arcsine terms. *Behavioural Brain Research*, **42**, 123–34.

Ringo, J. L. (1993). Spared short-term memory in monkeys following medial temporal lobe lesions is not yet established: a reply to Alvarez-Royo, Zola-Morgan and Squire. *Behavioural Brain Research*, **59**, 65–72.

Risold, P. Y. and Swanson, L. W. (1996). Structural evidence for functional domains in the rat hippocampus. *Science*, **272**, 1484–6.

Rivas, J., Gaztelu, J. M., and García-Austt, E. (1996). Changes in hippocampal cell discharge patterns and theta rhythm spectral properties as a function of walking velocity in the guinea pig. *Experimental Brain Research*, **108**, 113–18.

Robbins, T. W. (1996). Refining the taxonomy of memory. *Science*, **273**, 1353–4.

Roberts, A. C. (1996). Comparison of cognitive function in human and non-human primates. *Cognitive Brain Research*, **3**, 319–27.

Roberts, A. C., Robbins, T. W., and Everitt, B. J. (1988). The effects of intradimensional and extradimensional shifts on visual discrimination learning in humans and non-human primates. *Quarterly Journal of Experimental Psychology*, **40B**, 321–41.

Robertson, H. A., Martin, I. L., and Candy, J. M. (1978). Differences in benzodiazepine receptor binding in Maudsley reactive and Maudsley nonreactive rats. *European Journal of Pharmacology*, **50**, 455–7.

Robinson, D. S. (1993). Serotonin receptor subtypes and affective disorders. *Clinical Neuropharmacology*, **16**, (Suppl. 3), S1–S5.
Rodgers, R. J. and Cole, J. C. (1993). Influence of social isolation, gender, strain, and prior novelty on plus-maze behaviour in mice. *Physiology and Behavior*, **54**, 729–36.
Rodgers, R. J. and Cole, J. C. (1994). The elevated plus-maze: pharmacology, methodology and ethology. In *Ethology and Psychopharmacology* (ed. S. J. Cooper and C. A. Hendrie), pp. 9–44. Wiley.
Rogan, M. T., Stäubli, U. V., and LeDoux, J. E. (1997). Fear conditioning induces associative long-term potentiation in the amygdala. *Nature*, **390**, 604–7.
Roland, P. E. and Gulyas, B. (1994). Visual imagery and visual representation. *Trends in Neuroscience*, **17**, 281–6.
Rolls, E. T. and O'Mara, S. M. (1995). View-responsive neurons in the primate hippocampal complex. *Hippocampus*, **5**, 409–24.
Röper, G. and Rachman, S. (1976). Obsessional-compulsive checking: experimental replication and development. *Behaviour Research and Therapy*, **14**, 25–32.
Rose, G. and Day, S. (1990). The population mean predicts the number of deviant individuals. *British Medical Journal*, **391**, 1031–4.
Rosen, J. B., Hitchcock, J. M., Sananes, C. B., Miserendino, M. J. D., and Davis, M. (1991). A direct projection from the central nucleus of the amygdala to the acoustic startle pathway: anterograde and retrograde tracing studies. *Behavioral Neuroscience*, **105**, 817–25.
Rosenberg, R. and Jensen, P. N. (1994). Treatment of panic disorder with alprazolam and imipramine: clinical relevance of DSM III dysthymia. *European Journal of Psychiatry*, **9**, 27–32.
Rosenfeld, D. S. and Furman, Y. (1994). Pure sleep panic: two case reports and a review of the literature. *Sleep*, **17**, 462–5.
Rosenkilde, C. E. (1979). Functional heterogeneity of the prefrontal cortex in the monkey: a review. *Behavioral and Neural Biology*, **25**, 301–45.
Rosenzweig, M. R., Bennett, E. L., Colombo, P. J., Lee, D. W., and Serrano, P. A. (1993). Short-term, intermediate-term, and long-term memories. *Behavioural Brain Research*, **57**, 193–8.
Ross, R. T., Orr, W. B., Holland, P. C., and Berger, T. W. (1984). Hippocampectomy disrupts acquisition and retention of learned conditional responding. *Behavioral Neuroscience*, **98**, 211–25.
Rowan, M. J., Cullen, W. K., and Moulton, B. (1990). Buspirone impairment of performance of passive avoidance and spatial learning tasks in the rat. *Psychopharmacology (Berlin)*, **100**, 393–8.
Royce, J. R. (1977). On the construct validity of open-field measures. *Psychological Bulletin*, **84**, 1098–106.
Rudy, J. W. and Sutherland, R. J. (1989). The hippocampal formation is necessary for rats to learn and remember configural discriminations. *Behavioural Brain Research*, **34**, 97–109.
Rudy, J. W. and Sutherland, R. J. (1992). Wait a minute, Jarrard and Davidson's data support configural association theory. *Hippocampus*, **2**, 89
Rush, A. J., Beck, A. T., Kovacs, J. M., and Hollon, S. D. (1977). Comparative efficacy of cognitive therapy in outpatient depressives. *Cognitive Therapy and Research*, **1**, 17–37.
Sacco, W. P. and Beck, A. T. (1985). Cognitive therapy of depression. In *Handbook of depression: treatment, assessment and research* (ed. E. E. Beckham and W. R. Leber), pp. 3–38. The Dorsey Press, Homewood.

Sachdev, P. and Hay, P. (1996). Site and size of lesion and psychosurgical outcome in obsessive–compulsive disorder: a magnetic resonance imaging study. *Biological Psychiatry*, **39**, 739–42.

Sahgal, A. and Iversen, S. D. (1978). The effects of chlordiazepoxide on a delayed pair comparison task in pigeons. *Psychopharmacology (Berlin)*, **59**, 57–64.

Sainsbury, R. S. and Montoya, C. P. (1984). The relationship between type 2 theta and behavior. *Physiology and Behavior*, **33**, 621–6.

Samuels, I. and Valian, V. (1968). Hippocampal lesions and redundant visual cues: effects on spatial reversal learning and subsequent visual discriminations. *Proceedings of the 76th Annual Convention of the American Psychological Association*, **3**, 321–2.

Sanderson, W. C., Wetzler, S., and Asnis, G. M. (1994). Alprazolam blockade of CO_2-provoked panic in patients with panic disorder. *American Journal of Psychiatry*, **151**, 1220–2.

Sanger, D. J. (1985). GABA and the behavioral effects of anxiolytic drugs. *Life Sciences*, **36**, 1503–13.

Sanger, D. J., Benavides, J., Perrault, G., Morel, E., Cohen, C., Joly, D., *et al.* (1994). Recent developments in the behavioral pharmacology of benzodiazepine (omega) receptors: evidence for the functional significance of receptor subtypes. *Neuroscience and Biobehavioral Reviews*, **18**, 355–72.

Sapolsky, R. M., Uno, H., Rebert, C. S., and Finch, C. E. (1990). Hippocampal damage associated with prolonged glucocorticoid exposure in primates. *The Journal of Neuroscience*, **10**, 2897–902.

Sara, S. J., Devauges, V., and Biegon, A. (1993). Maudsley rat strains, selected for differences in emotional responses, differ in behavioral response to clonidine and in [^{125}I]clonidine binding in the locus coeruleus. *Behavioural Brain Research*, **57**, 101–4.

Sartory, G., MacDonald, R., and Gray, J. A. (1990). Effects of diazepam on approach, self-reported fear and psychophysiological responses in snake phobics. *Behaviour Research and Therapy*, **28**, 273–82.

Schallek, W., Schlosser, W., and Randall, L. O. (1972). Recent developments in the pharmacology of the benzodiazepines. *Advances in Pharmaceuticals for Chemotherapy*, **10**, 119–81.

Schuckit, M. A. (1984). Clinical studies of buspirone. *Psychopathology*, **17**, 61–8.

Schulkin, J. (1994). Melancholic depression and the hormones of adversity: a role for the amygdala. *Current Directions in Psychological Science*, **3**, 41–4.

Schulkin, J., McEwen, B. S., and Gold, P. W. (1994). Allostasis, amygdala, and anticipatory angst. *Neuroscience and Biobehavioral Reviews*, **18**, 385–96.

Schwegler, H. and Crusio, W. E. (1995). Correlations between radial-maze learning and structural variations of septum and hippocampus in rodents. *Behavioural Brain Research*, **67**, 29–41.

Schwegler, H. and Lipp, H.-P. (1983). Hereditary covariations of neuronal circuitry and behavior: correlations between the proportions of hippocampal synaptic fields in the regio inferior and two way avoidance in mice and rats. *Behavioural Brain Research*, 7, 1–38.

Scottish Office (1996). *Neurosurgery for mental disorders*. Her Majesty's Stationery Office, Edinburgh.

Seeley, R. J., Galaverna, O., Schulkin, J., Epstein, A. N., and Grill, H. J. (1993). Lesions of the central nucleus of the amygdala. II. Effects on intraoral NaCl intake. *Behavioural Brain Research*, **59**, 19–25.

Segal, M. (1977a). Excitability changes in rat hippocampus during conditioning. *Experimental Neurology*, **55**, 67–73.

Segal, M. (1977b). The effects of brainstem priming stimulation on interhemispheric hippocampal responses in the awake rat. *Experimental Brain Research*, **28**, 529–41.
Segal, M. and Bloom, F. E. (1974). The action of norepinephrine in the rat hippocampus. II. Activation of the input pathway. *Brain Research*, **72**, 99–114.
Segal, M. and Bloom, F. E. (1976). The action of norepinephrine in the rat hippocampus. IV. The effects of locus coeruleus stimulation on evoked hippocampal unit activity. *Brain Research*, **107**, 513–25.
Selden, N. R. W., Everitt, B. J., Jarrard, L. E., and Robbins, T. W. (1991). Complementary roles for the amygdala and hippocampus in aversive conditioning to explicit and contextual cues. *Neuroscience*, **42**, 335–50.
Selemon, L. D., Goldman-Rakic, P. S., and Tamminga, C. A. (1995). Prefrontal cortex. *American Journal of Psychiatry*, **152**, 5
Seligman, M. E. P. (1971). Phobias and preparedness. *The Behaviour Therapist*, **2**, 307–20.
Seligman, M. E. P. (1975). *Helplessness*. Freeman, San Francisco.
Shanks, D. R. and St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioural and Brain Sciences*, 367–95.
Sharp, P. E. and Green, C. (1994). Spatial correlates of firing patterns of single cells in the subiculum of the freely moving rat. *The Journal of Neuroscience*, **14**, 2339–56.
Sharp, P. E., Blair, H. T., Etkin, D., and Tzanetos, D. B. (1995). Influences of vestibular and visual motion information on the spatial firing patterns of hippocampal place cells. *The Journal of Neuroscience*, **15**, 173–89.
Shear, M. K. and Maser, J. D. (1994). Standardized assessment for panic disorder research: a conference report. *Archives of General Psychiatry*, **51**, 346–54.
Shear, M. K., Pilkonis, P. A., Cloitre, M., and Leon, A. C. (1994). Cognitive behavioral treatment compared with nonprescriptive treatment of panic disorder. *Archives of General Psychiatry*, **51**, 395–401.
Sheehan, D. V., Ballenger, J., and Jacobson, G. (1981). Relative efficacy of monoamine oxidase inhibitors and tricyclic antidepressants in the treatment of endogenous anxiety. In *Anxiety: new research and changing concepts* (ed. D. F. Klein and J. Rabkin), pp. 47–67. Raven Press, New York.
Sheehan, D. V., Raj, A. B., Sheehan, K. H., and Soto, S. (1990). Is buspirone effective for panic disorders? *Journal of Clinical Psychopharmacology*, **10**, 3–11.
Shekhar, A. (1994). Effects of treatment with imipramine and clonazepam on an animal model of panic disorder. *Biological Psychiatry*, **36**, 748–58.
Shibata, S., Yamashita, K., Yamamoto, E., Ozaki, T., and Ueki, S. (1989). Effects of benzodiazepine and GABA antagonists on anticonflict effects of antianxiety drugs injected into the rat amygdala in a water-lick suppression test. *Psychopharmacology (Berlin)*, **98**, 38–44.
Shin, L. M., Kosslyn, S. N., McNally, R. J., Alpert, N. M., Thompson, W. L., Rauch, S. L., *et al.* (1997). Visual imagery and perception in posttraumatic stress disorder. A positron emission tomographic investigation. *Archives of General Psychiatry*, **54**, 233–41.
Sidman, M., Stoddard, L., and Mohr, J. (1968). Some additional quantitative observations of immediate memory in a patient with bilateral hippocampal lesions. *Neuropsychologia*, **6**, 245–54.
Simon, P. and Soubrié, P. (1979). Behavior studies to differentiate anxiolytic and sedative activity of the tranquilizing drugs. In *Differential psychopharmacology of anxiolytics and sedatives. Modern problems of pharmacopsychiatry*, Vol. 14 (ed. J. R. Boissier), pp. 99–143. Karger, Basel.
Simonov, P. V. (1974). On the role of the hippocampus in the integrative activity of the brain. *Acta Neurobiologica Experimentalis*, **34**, 33–41.

Simons, A. D., Murphy, G. E., Levine, J. L., and Wetzel, R. D. (1986). Cognitive therapy and pharmacotherapy for depression: sustained improvement over one year. *Archives of General Psychiatry*, **43**, 43–8.

Sirigu, A., Zalla, T., Pillon, B., Grafman, J., Agid, Y., and Dubois, B. (1995). Selective impairments in managerial knowledge following pre-frontal cortex damage. *Cortex*, **31**, 301–16.

Skaggs, W. E., McNaughton, B. L., Wilson, M. A., and Barnes, C. A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus*, **6**, 149–72.

Skinner, B. F. (1938). *The behavior of organisms*. Appleton-Century, New York.

Skinner, D. M., Martin, G. M., Harley, C., Kolb, B., Pridgar, A., Bechara, A., *et al.* (1994). Acquisition of conditional discriminations in hippocampal lesioned and decorticated rats: evidence for learning that is separate from both simple classical conditioning and configural learning. *Behavioral Neuroscience*, **108**, 911–26.

Sledge, G. W., Einhorn, L., Nagy, C., and House, K. (1991). Phase III doubleblind comparison of intravenous ondansetron and metoclopromide as anti-emetic therapy for patients receiving multiple-day cisplatin-based chemotherapy. *American Journal of Psychiatry*, **136**, 1184–7.

Sluyter, F., Marican, C. C. M., and Crusio, W. E. (1999). Further phenotypical characterisation of two substrains of C57B1/6J inbred mice differing by a spontaneous single-gene mutation. *Behavioural Brain Research*, **98**, 39–43.

Smith, G. W., Aubry, J. M., Dellu, F., Contarino, A., Bilezikjian, L. M., Gold, L. H., *et al.* (1998). Corticotropin releasing factor receptor 1-deficient mice display decreased anxiety, impaired stress response, and aberrant neuroendocrine development. *Neuron*, **20**, 1093–102.

Smith, M. L. and Milner, B. (1984). Differential effects of frontal-lobe lesions on cognitive estimation and spatial memory. *Neuropsychologia*, **22**, 697–705.

Smythe, J. W., Murphy, D., Bhatnagar, S., Timothy, C., and Costall, B. (1996). Muscarinic antagonists are anxiogenic in rats tested in the black–white box. *Pharmacology, Biochemistry and Behavior*, **54**, 57–63.

Sokolov, E. N. (1960). Neuronal models and the orienting reflex. In *The central nervous system and behaviour, 3rd Conference* (ed. M. A. B. Brazier), pp. 187–276. Josiah Macy Jr Foundation, New York.

Sokolov, E. N. (1963). *Perception and the conditioned reflex*. Pergamon Press, Oxford.

Soldo, B. L., Proctor, W. R., and Dunwiddie, T. V. (1994). Ethanol differentially modulates $GABA_A$ receptor-mediated chloride currents in hippocampal, cortical, and septal neurons in rat brain slices. *Synapse*, **18**, 94–103.

Solomon, P. R. (1977). Role of the hippocampus in blocking and conditioned inhibition of the rabbit's nictating membrane response. *Journal of Comparative and Physiological Psychology*, **91**, 407–17.

Solomon, P. R. and Moore, J. W. (1975). Latent inhibition and stimulus generalization of the classically conditioned nictating membrane response in rabbits following dorsal hippocampal ablation. *Journal of Comparative and Physiological Psychology*, **89**, 1192–203.

Soubrié, P. (1986). Reconciling the role of central serotonin neurons in human and clinical behavior. *Behavioural and Brain Sciences*, **9**, 319.

Southwick, S. M., Bremner, D., Krystal, J. H., and Charney, D. S. (1994). Psychobiologic research in post-traumatic stress disorder. *Psychiatric Clinics of North America*, **17**, 251–64.

Spence, J. T. and Spence, K. W. (1966). The motivational components of manifest anxiety: drive and drive stimuli. In *Anxiety and behavior* (ed. C. D. Spielberger), pp. 291–326. Academic Press, London.

Squire, L. R. (1992). Declarative and nondeclarative memory: multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience*, **4**, 230–43.

Squire, L. R. and Cave, C. B. (1991). The hippocampus, memory and space. *Hippocampus*, **1**, 269–71.

Squire, L. R. and Zola-Morgan, S. (1984). The neuropsychology of memory: new links between humans and experimental animals. *Annals of the New York Academy of Sciences*, 137–49.

Squire, L. R. and Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, **253**, 1380–6.

Squire, L. R., Amaral, D. G., and Press, G. A. (1990). Magnetic resonance imaging of the hippocampal formation and mammillary nuclei distinguish medial temporal lobe and diencephalic amnesia. *The Journal of Neuroscience*, **10**, 3106–17.

Staddon, J. E. R. (1983). *Adaptive Behaviour and Learning*. Cambridge: Cambridge University Press.

Stampfl, T. G. (1970). Implosive therapy: an emphasis on covert stimulation. In *Learning approaches to therapeutic behavior change* (ed. D. J. Levis), pp. 182–204. Aldine, Chicago.

Steckler, T. and Muir, J. L. (1996). Measurement of cognitive function: relating rodent performance with human minds. *Cognitive Brain Research*, **3**, 299–308.

Stein, M. B., Asmundson, G. J. G., and Chartier, M. (1994a). Autonomic responsivity in generalized social phobia. *Journal of Affective Disorders*, **31**, 211–21.

Stein, M. B., Walker, J. R., and Forde, D. R. (1994b). Setting diagnostic thresholds for social phobia: considerations from a community survey of social anxiety. *American Journal of Psychiatry*, **151**, 408–12.

Stein, M. B., Koverola, C., Hanna, C., Torchia, M. G., and McClarty, B. (1997). Hippocampal volume in women victimized by childhood sexual abuse. *Psychological Medicine*, **27**, 951–9.

Steinbusch, H. W. M. (1981). Distribution of serotonin-immunoreactivity in the central nervous system of the rat—cell bodies and terminals. *Neuroscience*, **6**, 557–618.

Stenzel-Poore, M. P., Heinrichs, S. C., Rivest, S., Koob, G. F., and Vale, W. W. (1994). Overproduction of corticotropin-releasing factor in transgenic mice: a genetic model of anxiogenic behavior. *The Journal of Neuroscience*, **14**, 2579–84.

Stern, R. and Marks, I. M. (1973). Brief and prolonged flooding: a comparison in agoraphobic patients. *Archives of General Psychiatry*, **28**, 270–6.

Ström-Olsen, R. and Carlisle, R. (1971). Bi-frontal stereotactic tractotomy: a follow-up study of its effects on 210 patients. *British Journal of Psychiatry*, **118**, 141–54.

Sudré, E. C. M., De Barros, M. R., Sudré, G. N., and Schenberg, L. C. (1993). Thresholds of electrically induced defence reaction of the rat: short- and long-term adaptation mechanisms. *Behavioural Brain Research*, **58**, 141–54.

Sue, D. (1975). The effect of duration of exposure on systematic desensitization and extinction. *Behaviour Research and Therapy*, **13**, 55–60.

Sutherland, N. S. and Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning*. Academic Press, London.

Sutherland, R. J. and Rudy, J. W. (1989). Configural association theory: the role of the hippocampal formation in learning, memory and amnesia. *Psychobiology*, **17**, 129–44.

Suzuki, W. A. and Amaral, D. G. (1994). Topographic organization of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices. *The Journal of Neuroscience*, **14**, 1856–77.

Suzuki, W. A., Zola-Morgan, S., Squire, L. R., and Amaral, D. G. (1993). Lesions of the perirhinal and parahippocampal cortices in the monkey produce long-lasting memory impairment in the visual and tactual modalities. *The Journal of Neuroscience*, **13**, 2430–51.

Swanson, L. W. (1978). The anatomical organization of septo-hippocampal projections. In *Functions of the septo-hippocampal system*, Ciba Foundation Symposium 58 (New Series) (ed. K. Elliott and J. Whelan), pp. 25–43. Elsevier, Amsterdam.

Swanson, L. W. (1995). Mapping the human brain: past, present, and future. *Trends in Neuroscience*, **18**, 471–4.

Swayze, V. W., II. (1995). Frontal leukotomy and related psychosurgical procedures in the era before antipsychotics (1935–1954): A historical overview. *American Journal of Psychiatry*, **152**, 505–15.

Swoboda, K. J. and Jenike, M. A. (1995). Frontal abnormalities in a patient with obsessive–compulsive disorder: the role of structural lesions in obsessive–compulsive behavior. *Neurology*, **45**, 2130–4.

Tallman, J. F., Paul, S. M., Skolnick, P., and Gallager, D. W. (1980). Receptors for the age of anxiety: pharmacology of the benzodiazepines. *Science (NY)*, **267**, 274–81.

Talmi, M., Carlier, E., Bengelloun, W., and Soumireu-Mourat, B. (1995). Synergistic action of corticosterone on kainic acid-induced electrophysiological alterations in the hippocampus. *Brain Research*, **704**, 97–102.

Tan, E., Marks, I. M., and Marset, P. (1971). Bimedial leucotomy in obsessive compulsive neurosis: a controlled serial enquiry. *British Journal of Psychiatry*, **118**, 155–64.

Tan, S., Kirk, R. C., Abraham, W. C., and McNaughton, N. (1996). Chlordiazepoxide reduces discriminability but not rate of forgetting in delayed conditional discrimination. *Psychopharmacology (Berlin)*, **101**, 550–4.

Tanaka, D. and Goldman, P. S. (1976). Silver degeneration and autoradiographic evidence for a projection from the principal sulcus to the septum in the Rhesus monkey. *Brain Research*, **103**, 535–40.

Tanila, H., Shapiro, M. L., and Eichenbaum, H. (1997). Discordance of spatial representation in ensembles of hippocampal place cells. *Hippocampus*, 7, 613–23.

Taub, J., Taylor, P., Smith, M., Kelley, K., Becker, B., and Reid, L. (1977). Methods of deconditioning persisting avoidance: drugs as adjuncts to response prevention. *Physiological Psychology*, **5**, 67–72.

Taube, J. S., Muller, R. U., and Ranck, J. B., Jr. (1990a). Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *The Journal of Neuroscience*, **10**, 420–35.

Taube, J. S., Muller, R. U., and Ranck, J. B., Jr. (1990b). Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations. *The Journal of Neuroscience*, **10**, 436–47.

Taube, J. S., Goodridge, J. P., Golob, E. J., Dudchenko, P. A., and Stackman, R. W. (1996). Processing the head direction cell signal: a review and commentary. *Brain Research Bulletin*, **40**, 477–86.

Taylor, J. A. (1953). A personality scale of manifest anxiety. *Journal of Abnormal Social Psychology*, **48**, 285–90.

Teasdale, J. D. (1977). Psychological treatment of phobias. In *Tutorial essays in psychology*, Vol. 2 (ed. N. S. Sutherland), pp. 137–63. Erlbaum, Hillsdale, NJ.

Teicher, M. H. (1988). Biology of anxiety. *Medical Clinics of North America*, **72**, 791–814.

Teyler, T. and Discenna, P. (1986). The hippocampal memory indexing theory. *Behavioral Neuroscience*, **100**, 147–54.

Thaler, D., Chen, Y.-C., Nixon, P. D., Stern, C. E., and Passingham, R. E. (1995). The functions of the medial premotor cortex. I. Simple learned movements. *Experimental Brain Research*, **102**, 445–60.

Thomas, G. J. (1978). Delayed alternation in rats after pre- or postcommissural fornicotomy. *Journal of Comparative and Physiological Psychology*, **92**, 1128–36.

Thomas, R. K. (1996). Investigating cognitive abilities in animals: unrealized potential. *Cognitive Brain Research*, **3**, 157–66.

Ticku, M. K. and Olsen, R. W. (1978). Interaction of barbiturates with dihydropicrotoxinin binding sites related to the GABA receptor-ionophore system. *Life Sciences*, **22**, 1643–52.

Toates, F. (1995). *Stress: conceptual and biological aspects*. Wiley, Chichester.

Toczek, S. (1960). Disturbances of inhibitory processes due to lesions in the frontal area of the brain in man. *Acta Biologica Experimentalis*, **20**, 103–19.

Tomaz, C., Dickinson-Anson, H., McGaugh, J. L., Souza-Silva, M. A., Viana, M. B., and Graeff, F. G. (1993). Localization in the amygdala of the amnestic action of diazepam on emotional memory. *Behavioural Brain Research*, **58**, 99–105.

Torgersen, S. (1979). The nature and origin of common phobic fears. *British Journal of Psychiatry*, **134**, 343–51.

Tranel, D. and Hyman, T. (1990). Neuropsychological correlates of bilateral amygdala damage. *Archives of Neurology*, **47**, 349–55.

Treit, D., Pesold, C., and Rotzinger, S. (1993). Noninteractive effects of diazepam and amygdaloid lesions in two animal models of anxiety. *Behavioral Neuroscience*, **107**, 1099–105.

Tremblay, M., Lacroix, D., Chaput, Y., Fraïle, V., Lamer, R., and Albert, J.-M. (1994). Brain activation with a maze test: an EEG coherence analysis study in healthy subjects. *NeuroReport*, **5**, 2449–53.

Trimble, M. R. (1990). Benzodiazepines in clinical practice. In *The anxiolytic jungle: where next* (ed. D. Wheatley), pp. 9–21. Wiley, Chichester.

Tripp, G. and McNaughton, N. (1987). Naloxone fails to block the effects of chlordiazepoxide on acquisition and performance of successive discrimination. *Psychopharmacology*, **91**, 119–23.

Tripp, G. and McNaughton, N. (1992). Schedule dependence of the interaction of naloxone and chlordiazepoxide. *Pharmacology, Biochemistry and Behavior*, **41**, 475–81.

Tripp, G., McNaughton, N., and Oei, T. P. S. (1987). Naloxone blocks the effects of chlordiazepoxide on acquisition but not performance of differential reinforcement of low rates of response (DRL). *Psychopharmacology (Berlin)*, **91**, 112–18.

Tulving, E. (1984). Multiple book review of *Elements of episodic memory*. *Behavioural and Brain Sciences*, **7**, 223–38.

Tulving, E. (1985). On the classification problem in learning and memory. In *Perspectives on learning and memory* (ed. L.-G. Nilsson and T. Archer), pp. 67–94. Erlbaum, Hillsdale, NJ.

Urban, J. H., Van de Kar, L. D., Lorens, S. A., and Bethea, C. L. (1986). Effect of the anxiolytic drug buspirone on prolactin and corticosterone secretion in stressed and unstressed rat. *Pharmacology, Biochemistry and Behavior*, **25**, 457–62.

Ursin, H. (1980). Affective and instrumental aspects of fear and aggression. In *Functional states of the brain: their determinants* (ed. M. Koukkou, D. Lehmann, and J. Agnst), pp. 119–30. Elsevier.

Ursin, H. (1985). The instrumental effects of emotional behavior. In *Perspectives in ethology* (ed. P. P. G. Bateson and P. H. Klopfer), pp. 45–62. Plenum.

Valentine, S. (1930). The innate basis of fear. *Journal of Genetic Psychology*, **37**, 394–419.
Valero, I., Stewart, J., McNaughton, N., and Gray, J. A. (1977). Septal driving of the hippocampal theta rhythm as a function of frequency in the male rat: effects of adrenopituitary hormones. *Neuroscience*, **2**, 1029–32.
Van Ameringen, M., Mancini, C., and Streiner, D. (1994). Sertraline in social phobia. *Journal of Affective Disorders*, **31**, 141–5.
Van Daal, J. H. H. M., Herbergs, P. J., Cursion, W. E., Schwegler, H., Jenks, B. G., Lemmens, W. A. J. G., *et al.* (1991a). A genetic-correlational study of hippocampal structural variation and variation in exploratory activities of mice. *Behavioural Brain Research*, **43**, 57–64.
Van Daal, J. H. H. M., Jenks, B. G., Crusio, W. E., Lemmens, W. A. J. G., and van Abeelen, J. H. F. (1991b). A genetic-correlational study of hippocampal neurochemical variation and variation in exploratory activities of mice. *Behavioural Brain Research*, **43**, 65–72.
Van der Kolk, B. A., Greenberg, M. S., Orr, S. P., and Pitman, R. K. (1989). Endogenous opioids, stress induced analgesia, and posttraumatic stress disorder. *Psychopharmacology Bulletin*, **25**, 417–21.
Van Hoesen, G. W. and Pandya, D. N. (1975). Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the Rhesus monkey. I. Temporal lobe afferents. *Brain Research*, **95**, 1–24.
Van Hoesen, G. W., Pandya, D. N., and Butters, N. (1972). Cortical afferents to the entorhinal cortex of the Rhesus monkey. *Science (NY)*, **175**, 1471–3.
Van Hoesen, G. W., Pandya, D. N., and Butters, N. (1975). Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the Rhesus monkey. II. Frontal lobe afferents. *Brain Research*, **95**, 25–38.
Van Hoesen, G. W., Rosene, D. L., and Mesulam, M. M. (1979). Subicular input from temporal cortex in the Rhesus monkey. *Science (NY)*, **205**, 608–10.
Vanderwolf, C. H. (1969). Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalography and Clinical Neurophysiology*, **26**, 407–18.
Vanderwolf, C. H. (1995). Does a history of convulsions increase the amnestic effect of temporal region brain lesions? *Physiology and Behavior*, **57**, 193–7.
Veltman, D. J., Van Zijderveld, G. A., and Van Dyck, R. (1996). Epinephrine infusions in panic disorder: a double-blind placebo-controlled study. *Journal of Affective Disorders*, **39**, 133–40.
Verbanac, J. S., Commissaris, R. L., Altman, H. J., and Pitts, D. K. (1994). Electrophysiological characteristics of locus coeruleus neurons in the Maudsley reactive (MR) and non-reactive (MNRA) rat strains. *Neuroscience Letters*, **179**, 137–40.
Verbanac, J. S., Commissaris, R. L., and Pitts, D. K. (1995). An electrophysiological evaluation of serotonergic dorsal raphe neurons in Maudsley rats. *Life Sciences*, **58**, 245–50.
Vertes, R. P. and Kocsis, B. (1997). Brainstem–diencephalo-septohippocampal systems controlling the theta rhythm of the hippocampus. *Neuroscience*, **81**, 893–926.
Viallet, F., Vuillon-Cacciuttolo, G., Legallet, E., Bonnefoi-Kyriacou, B., and Trouche, E. (1995). Bilateral and side-related reaction time impairments in patients with unilateral cerebral lesions of a medial frontal region involving the supplementary motor area. *Neuropsychologia*, **33**, 215–23.
Viana, M. B., Tomaz, C., and Graeff, F. G. (1994). The elevated T-maze: a new animal model of anxiety and memory. *Pharmacology, Biochemistry and Behavior*, **49**, 549–54.
Victor, M., Adams, R. D., and Collins, G. H. (1971). *The Wernicke–Korsakoff syndrome*. Blackwell, Oxford.

Vidailhet, P., Danion, J.-M., Kauffmann-Muller, F., Grangé, D., Giersch, A., Van Der Linden, M., *et al.* (1994). Lorazepam and diazepam effects on memory acquisition in priming tasks. *Psychopharmacology (Berlin)*, **115**, 397–406.

Vinogradova, O. S. (1975). Functional organization of the limbic system in the process of registration of information: facts and hypotheses. In *The Hippocampus*, Vol. 2 (ed. R. I. Isaacson and K. H. Pribram), pp. 3–69. Plenum Press, New York.

Vinogradova, O. S. and Brazhnik, P. S. (1978). Neuronal aspects of septo-hippocampal relations. In *Functions of the septo-hippocampal system*, Ciba Foundation Symposium 58 (New Series) (ed. K. Elliott and J. Whelan), pp. 145–71. Elsevier, Amsterdam.

Virley, D., Ridley, R. M., Sinden, J. D., Kershaw, T. R., Harland, S., Rashid, T., *et al.* (1999). Primary CA1 and conditionally immortal MHP36 cell grafts restore conditional discrimination learning and recall in marmosets after excitotoxic lesions of the hippocampal CA1 field. *Brain*, **122**, 2321–335.

Vnek, N. and Rothblat, L. A. (1996). The hippocampus and long-term object memory in the rat. *The Journal of Neuroscience*, **16**, 2780–7.

Vogt, B. A. (1985). Cingulate cortex. In *Cerebral cortex*: *association and auditory cortices* (ed. A. Peters and E. G. Jones), pp. 89–149. Plenum Press, New York.

Wagner, A. R. (1963). Conditioned frustration as a learned drive. *Journal of Experimental Child Psychology*, **66**, 142–8.

Wagner, A. R. (1966). Frustration and punishment. In *Current research on motivation* (ed. R. M. Haber), pp. 229–39. Holt, Rinehart, Winston, New York.

Wagner, A. R. (1969). Frustrative nonreward: a variety of punishment? In *Punishment and aversive behavior* (ed. B. A. Campbell and R. M. Church), pp. 157–81. Appleton-Century-Crofts, New York.

Wagner, A. R. (1978). Experiences and the priming of STM. In *Cognitive Processes in Animal Behaviour* (ed. S. H. Hulse, H. Fowler, and W. K. Honig), pp. 177–209. Erlbaum, Hillsdale, NJ.

Walker, J. A. and Olton, D. S. (1979). Spatial memory deficit following fimbria-fornix lesions: independent of time for stimulus processing. *Physiology and Behavior*, **23**, 11–15.

Wan, R.-Q., Pang, K., and Olton, D. S. (1994). Hippocampal and amygdaloid involvement in nonspatial and spatial working memory in rats: effects of delay and interference. *Behavioral Neuroscience*, **108**, 866–82.

Wardle, J., Hayward, P., Higgit, A., Stabl, M., Blizard, R., and Gray, J. A. (1994). Effects of concurrent diazepam treatment on the outcome of exposure therapy on agoraphobia. *Behaviour Research and Therapy*, **32**, 203–15.

Warrington, E. K. and Weiskrantz, L. (1968). A study of learning and retention in amnesic patients. *Neuropsychologia*, **6**, 283–91.

Warrington, E. K. and Weiskrantz, L. (1970). Amnesic syndrome: consolidation or retrieval? *Nature (London)*, **228**, 628–30.

Warrington, E. K. and Weiskrantz, L. (1974). The effect of prior learning on subsequent retention in amnesic patients. *Neuropsychologia*, **12**, 419–28.

Warrington, E. K. and Weiskrantz, L. (1978). Further analysis of the prior learning effect in amnesic patients. *Neuropsychologia*, **16**, 169–77.

Watanabe, M. (1996). Reward expectancy in primate prefrontal neurons. *Nature*, **382**, 629–32.

Watson, D., Clark, L. A., and Harkness, A. R. (1994). Structures of personality and their relevance to psychopathology. *Journal of Abnormal Psychology*, **103**, 18–31.

Watson, J. B. and Rayner, B. (1920). Conditioned emotional reactions. *Journal of Experimental Child Psychology*, **3**, 1–14.

Watts, F. N. (1971). Desensitization as an habituation phenomenon: stimulus intensity as determinant of the effect of stimulus lengths. *Behaviour Research and Therapy*, **9**, 209–17.

Watts, F. N. (1973). Desensitization as an habituation phenomenon. II. Studies of inter-stimulus interval length. *Psychological Reports*, **33**, 715–18.

Watts, F. N. (1974). The control of spontaneous recovery of anxiety in imaginal desensitization. *Behaviour Research and Therapy*, **12**, 57–9.

Watts, F. N. (1979). Habituation model of systematic desensitization. *Psychological Bulletin*, **86**, 627–37.

Waziri, R. (1990). Psychosurgery for anxiety and obsessive–compulsive disorders. In *Handbook of Anxiety*, Vol. 4, *The treatment of anxiety* (ed. R. Noyes, M. Roth, and G. D. Burrows), pp. 519–35. Elsevier, Amsterdam.

Webster, D. B. and Voneida, T. J. (1964). Learning deficits following hippocampal lesions in split-brain cats. *Experimental Neurology*, **10**, 170–82.

Weess, H. G., Steinberg, R., and Pritzel, M. (1994). Mnestic disorders in long-term high-dose diazepam abuse: a pilot study. *European Journal of Psychiatry*, **9**, 254–9.

Weiskrantz, L. (1978). A comparison of hippocampal pathology in man and other animals. In *Functions of the septo-hippocampal system*, Ciba Foundation Symposium 58 (New Series) (ed. K. Elliott and J. Whelan), pp. 373–87. Elsevier, Amsterdam.

Weiskrantz, L. and Warrington, E. K. (1970a). A study of forgetting in amnesic patients. *Neuropsychologia*, **8**, 281–8.

Weiskrantz, L. and Warrington, E. K. (1970b). Verbal learning and retention by amnesic patients using partial information. *Psychonomic Science*, 20, 210–11.

Weiskrantz, L. and Warrington, E. K. (1975). The problem of the amnesic syndrome in man and animals. In *The hippocampus*, Vol. 2, *Neurophysiology and behaviour* (ed. R. L. Isaacson and K. H. Pribram), pp. 411–28. Plenum Press, New York.

Weiskrantz, L. and Warrington, E. K. (1979). Conditioning in amnesic patients. *Neuropsychologia*, **17**, 187–94.

Weiss, C., Kronforst-Collins, M. A., and Disterhoft, J. F. (1996). Activity of hippocampal pyramidal neurons during trace eyeblink conditioning. *Hippocampus*, **6**, 192–209.

Weiss, K. R., Friedman, R., and McGregor, S. (1974). Effects of septal lesions on latent inhibition and habituation of the orienting response in rats. *Acta Neurobiologicae Experimentalis*, **34**, 491–504.

Whishaw, I. Q. and Tomie, J. (1991). Acquisition and retention by hippocampal rats of simple, conditional, and configural tasks using tactile and olfactory cues: implications for hippocampal function. *Behavioral Neuroscience*, **105**, 787–97.

Whishaw, I. Q. and Tomie, J.-A. (1995). Rats with fimbria–fornix lesions can acquire and retain a visual–tactile transwitching (configural) task. *Behavioral Neuroscience*, **109**, 607–12.

Whishaw, I. Q. and Tomie, J.-A. (1997). Perseveration on place reversals in spatial swimming pool tasks: further evidence for place learning in hippocampal rats. *Hippocampus*, 7, 361–70.

Whishaw, I. Q. and Vanderwolf, C. H. (1973). Hippocampal EEG and behavior: changes in amplitude and frequency of RSA (theta rhythm) associated with spontaneous and learned movement patterns in rats and cats. *Behavioural Biology*, **8**, 461–84.

Whishaw, I. Q., Cassel, J.-C., and Jarrard, L. E. (1995). Rats with fimbria–fornix lesions display a place response in a swimming pool: a dissociation between getting there and knowing where. *The Journal of Neuroscience*, **15**, 5779–88.

White, K. G. (1985). Characteristics of forgetting functions in delayed matching to sample. *Journal of the Experimental Analysis of Behaviour*, **44**, 15–34.
White, K. G. and Harper, D. N. (1996). Quantitative reanalysis of lesion effects on rate of forgetting in macaques. *Behavioural Brain Research*, **74**, 223–7.
White, K. G., Ruske, A. C., and Colombo, M. (1996). Memory procedures, performance and processes in pigeons. *Brain Research*, **16**, 1–9.
Wickens, D. D. (1970). Encoding categories of words: an empirical approach to meaning. *Psychological Review*, 77, 1–15.
Wiener, S. I. (1996). Spatial, behavioral and sensory correlates of hippocampal CA1 complex spike cell activity: Implications for information processing functions. *Progress in Neurobiology*, **49**, 335–61.
Wiener, S. I. and Korshunov, V. A. (1996). Place-independent behavioural correlates of hippocampal neurones in rats. *NeuroReport*, 7, 183–8.
Wikinski, S., Lombardo, M., Medina, J. H., and Rubio, M. C. (1994). Lack of anxiolytic effect of diazepam in pre-anaesthetic medication. *British Journal of Anaesthesiology*, **72**, 694–6.
Wilcock, J. and Fulker, D. W. (1973). Avoidance learning in rats: genetic evidence for two distinct behavioral processes in the shuttle-box. *Journal of Comparative and Physiological Psychology*, **82**, 247–53.
Willett, R. (1960). The effects of psychosurgical procedures on behaviour. In *Handbook of abnormal psychology* (ed. H. J. Eysenck), pp. 566–610. Pitman, London.
Williams, J., Gray, J. A., Snape, M., and Holt, L. (1989). Long-term effects of septohippocampal stimulation on behavioural responses to anxiogenic stimuli. In *Psychopharmacology of anxiety* (ed. P. Tyrer), pp. 80–108. Oxford University Press.
Williamson, M. J., Paul, S. M., and Skolnick, P. (1978). Labelling of benzodiazepine receptors in vivo. *Nature (London)*, **275**, 551–3.
Willner, P. J. and Crowe, R. (1977). Effect of chlordiazepoxide on the partial reinforcement extinction effect. *Pharmacology, Biochemistry and Behavior*, 7, 4779–82.
Wilson, M. A. and McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science*, **261**, 1055–8.
Wilson, M. A. and McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, **265**, 676–9.
Winocur, G. (1979). Effects of interference on discrimination learning and recall by rats with hippocampal lesions. *Physiology and Behavior*, **22**, 339–45.
Winocur, G. (1981). The amnesic syndrome: a deficit in cue utilization. In *Human memory and amnesia* (ed. L. Cermak), pp. 139–69. Erbaum, Hillsdale, NJ.
Winocur, G. (1985). The hippocampus and thalamus: their roles in short- and long-term memory and the effects of interference. *Behavioural Brain Research*, **16**, 135–52.
Winocur, G. (1992). A comparison of normal old rats and young adult rats with lesions to the hippocampus or prefrontal cortex on a test of matching-to-sample. *Neuropsychologia*, **30**, 769–81.
Winocur, G. and Kinsbourne, M. (1978). Contextual cueing as an aid to Korsakoff amnesics. *Neuropsychologia*, **16**, 671–82.
Winocur, G. and Mills, J. A. (1970). Transfer between related and unrelated problems following hippocampal lesions in rats. *Journal of Comparative and Physiological Psychology*, **73**, 162–9.
Winocur, G. and Olds, J. (1978). Effects of context manipulation on memory and reversal learning in rats with hippocampal lesions. *Journal of Comparative and Physiological Psychology*, **92**, 312–21.
Winocur, G. and Weiskrantz, L. (1976). An investigation of paired-associate learning in amnesic patients. *Neuropsychologia*, **14**, 97–110.

Winter, J. C. and Petti, D. T. (1987). The effects of 8-hydroxy-2-(di-*N*-propylamino)tetralin and other serotonergic agonists on performance in a radial maze: a possible role for $5HT_{1A}$ receptors in memory. *Pharmacology, Biochemistry and Behavior*, **27**, 625–8.

Wittchen, H.-U., Zhao, S., Kessler, R. C., and Eaton, W. W. (1994). DSM-III-R generalized anxiety disorder in the National Comorbidity Survey. *Archives of General Psychiatry*, **51**, 355–64.

Wolpe, J. (1958). *Psychotherapy by reciprocal inhibition*. Stanford University Press.

World Health Organization (1992). *The ICD-10 Classification of Mental and Behavioural Disorders*. World Health Organization, Geneva.

Wright, A. A. (1994). Primacy effects in animal memory and human nonverbal memory. *Animal Learning and Behaviour*, **22**, 219–23.

Yadin, E., Thomas, E., Strickland, C. E., and Grishkat, H. L. (1991). Anxiolytic effects of benzodiazepines in amygdala-lesioned rats. *Psychopharmacology (Berlin)*, **103**, 473–9.

Yamamoto, T., Shimura, T., Sako, N., Yasoshima, Y., and Sakai, N. (1994). Neural substrates for conditioned taste aversion in the rat. *Behavioural Brain Research*, **65**, 123–37.

Yee, B. K. and Rawlins, J. N. P. (1994). The effects of hippocampal formation ablation or fimbria–fornix section on performance of a nonspatial radial arm maze task by rats. *The Journal of Neuroscience*, **14**, 3766–74.

Yeragani, V. K., Srinivasan, K., Balon, R., Ramesh, C., and Berchou, R. (1994). Lactate sensitivity and cardiac cholinergic function in panic disorder. *American Journal of Psychiatry*, **151**, 1226–8.

Young, B. J. and McNaughton, N. (1997). Activity of hippocampal neurons in rats during DRL performance. *Society for Neuroscience Abstracts*, **23**, 503.

Zanotti, A., Arban, R., Perazzolo, M., and Giusti, P. (1994). Diazepam impairs place learning in naive but not in maze-experienced rats in the Morris water maze. *Psychopharmacology (Berlin)*, **115**, 73–8.

Zhu, X. and McNaughton, N. (1994). A comparison of the acute effects of a tricyclic and a MAOI antidepressant on septal driving of hippocampal rhythmical slow activity. *Psychopharmacology (Berlin)*, **114**, 337–44.

Zhu, X. O. and McNaughton, N. (1995). Similar effects of buspirone and chlordiazepoxide on a fixed interval schedule with long-term, low-dose administration. *Journal of Psychopharmacology*, **9**, 326–30.

Zinbarg, R. E. (1990). Animal research and behaviour therapy. Part I. Behaviour therapy is not what you think it is. *The Behaviour Therapist*, **13**, 171–5.

Zinbarg, R. E. and Barlow, D. H. (1996). Structure of anxiety and the anxiety disorders: a hierarchical model. *Journal of Abnormal Psychology*, **105**, 181–93.

Zinbarg, R. E. and Mineka, S. (1991). Animal models of psychopathology. II. Simple phobia. *The Behaviour Therapist*, 61–5.

Zola-Morgan, S., Squire, L. R., and Amaral, D. G. (1986). Human amnesia and the medial temporal region: enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus. *The Journal of Neuroscience*, **6**, 2950–67.

Zola-Morgan, S., Squire, L. R., Amaral, D. G., and Suzuki, W. A. (1989). Lesions of perirhinal and parahippocampal cortex that spare the amygdala and hippocampal formation produce severe memory impairment. *The Journal of Neuroscience*, **9**, 4355–70.

Zola-Morgan, S., Squire, L. R., and Ramus, S. J. (1994). Severity of memory impairment in monkeys as a function of locus and extent of damage within the medial temporal lobe memory system. *Hippocampus*, **4**, 483–95.

Index